DOCUMENT RESUME

ED 310 128 TM 013 701

AUTHOR Eggen, T. J. H. M., Ed.
TITLE Open Session IEA General Assembly (24th, Enschede, Netherlands, August 15-19, 1983).
INSTITUTION International Association for the Evaluation of Educational Achievement.; Twente Univ., Enschede (Netherlands). Dept. of Education.
PUB DATE Dec 83
NOTE 110p.
PUB TYPE Collected Works - Conference Proceedings (021) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Academic Achievement; *Classroom Environment; *Educational Assessment; Educational Policy; Elementary Secondary Education; Equal Education; Foreign Countries; International Organizations; International Programs; *International Studies; *Mathematics Achievement; National Programs; Outcomes of Education; *Science Education; Testing Programs; *Writing Evaluation
IDENTIFIERS IEA Written Composition Study; *International Assn Evaluation Educ Achievement; Second International Mathematics Study; Second International Science Study

ABSTRACT

Ten papers, four general overviews,, and three commentaries delivered at the General Assembly of the International Association for the Evaluation of Educational Achievement (IEA) in 1983 are presented. The papers include: (1) "Why Join IEA?" (J. P. Keeves); (2) "Research and Policymaking in Education: An International Perspective" (T. Husen); and (3) "A Diagnostic Way of Handling the Test-Curriculum Overlap Using Constrained Multidimensional Scaling" (W. de Corte and C. Brusselmans-Dehairs). A general overview by R. W. Phillipps of the Second International Mathematics Study precedes the fourth paper: (4) "Some Results of the Second International Mathematics Study in The Netherlands" (T. J. Eggen et al.). A general overview by A. Purves and S. Takala of the IEA Written Composition study is followed by the fifth paper: (5) "Results and Effects of IEA Written Composition Study in The Netherlands" (H. Wesdorp). A general overview by B. Avalos of the Classroom Environment Study precedes the sixth paper: (6) "Student Activities and Learning Outcomes" (W. Tomic and E. Warries). A general overview by J. P. Keeves of the Second IEA Science Study precedes the seventh paper: (7) "Optimalization of Reporting Results from National Assessment Studies" (W. J. Pelgrum). Concerning the issue of equality in educational opportunity, the following papers were delivered: (8) "Schooling and Equality" (J. S. Coleman); (9) "Phases in Social Structure and Change of Educational Opportunity, A Comment on Coleman's Paper" (J. Dronkers); and (10) "Designing a Policy for Equality of Educational Opportunity, A Comment on Coleman's Paper" (A. Hoogerwerf). Commentaries include the papers delivered by J. Dronkers and A. Hoogerwerf as well as commentaries by R. W. Phillipps, A. Purves, and J. P. Keeves, respectively, on papers on the international mathematics, composition, and science studies. (TJH)

ED310128

Twente University of Technology

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. NELISSEN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

OPEN SESSION

GENERAL ASSEMBLY 1983

Redactie: T.J.H.M. Eggen

TO Department of Education

BEST COPY AVAILABLE

COLOPHON

Title: Open Session IEA General Assembly 1983

Editor: T.J.H.M. Eggen

Publisher: Twente University of Technology, Department of Education

Typing: Secretaries Dept. of Education

Edition: 200

Date: December 1983.

CONTENTS

Preface

SECTION I:

J.P. Keeves: Why join IEA? 1

T. Husén: Research and policymaking in education: An international perspective. 5

W. de Corte, C. Brusselmans-Dehairs: A diagnostic way of handling the Test-Curriculum overlap using Constrained Nonmetric Multidimensional Scaling. 18

SECTION II:

The Second International Mathematics Study.

R.W. Phillipps: General Overview 29

T.J.H.M. Eggen, W.J. Pelgrum, Tj. Plomp: Some results of the second international mathematics study in The Netherlands. 31

R.W. Phillipps: Comment on the Dutch paper. 44

SECTION III:

The International Study of Achievement in Written Composition.

A. Purves, S. Takula: General Overview* 50

H. Wesdorp: Results and effects of IEA Written Compositon Study in The Netherlands. 52

A. Purves: Comment on the Dutch paper. 58

SECTION IV:

The Classroom Environment Study.

B. Avalos: General Overview** 60

W. Tomic, E. Warries: Student activities and learning outcomes. 62

SECTION V:

The Second IEA Science Study.

J.P. Keeves: General Overview. 72

W.J. Pelgrum: Optimalization of reporting results from national Assesment Studies. 74

J.P. Keeves: Comment on the Dutch paper. 82

SECTION VI:

Schooling and Equality.

J.S. Coleman:	Schooling and Equality.	85
J. Dronkers:	Phases in social structure and change of educational opportunity, a comment on Coleman's paper.	91
A. Hoogerwerf:	Designing a policy for equality of educational opportunity, a comment on Coleman's paper.	93

*Reprinted from IEA-newsletter, no. 2, july 1982.

**Reprinted from IEA-newsletter, no. 1, january 1982.

Preface

The Department of Education of Twente University of Technology in Enschede, The Netherlands hosted the 24th General Assembly of the International Association for the Evaluation of Educational Achievement (IEA) from 15 to 19 August 1983.
IEA is a cooperative organization of educational research centers in more than forty countries which co-operate in conducting cross-national empirical educational research.
During the General Assembly, which takes place annually, representatives of the member research institutions discuss and make decisions on the conduct and financing of current and future IEA research projects.

Since 1980 it has been customary for IEA to organize an open session during the General Assembly. During this open session educational researchers and the possible customers of the results of educational research are given the opportunity of obtaining more information about IEA and its research projects. The papers given during the open session have been collected into a report.
The present report consists of 6 sections.
In section I there are three papers: The first is by J.P. Keeves who discussed the benefits of participation of countries in IEA-research. T. Husén's paper is a comparative study on how research and policy making relate to each other in four countries: Sweden, the Federal Republic of Germany, Great Britain and the United States. The final paper in section I is a methodological contribution from W. de Corte and C. Brusselmans, who explore the use of a special multi-dimensional scaling technique for the overlap between tests and curriculum.
Section II to V all have the same structure. In these sections Dutch researchers in IEA-projects present some of the results (or plans) of their projects in The Netherlands. Each Dutch paper is preceded by a short general overview of the international project. At the end of each section there is a comment on the Dutch paper by the chairman of the international project council.
Successively the following are addressed: The Second International Mathematics Study, the International Study of Achievement in Written Composition, the Classroom Environment Study and the Second IEA Science Study.
Section VI of the report presents a new contribution to the discussion of the subject of schooling and equality. J.S. Coleman proposes a new perspective on the problem of equal educational opportunity based on comparisons between different societies. Coleman's paper is followed by invited comments from J. Dronkers and A. Hoogerwerf.

I hope that the publication of this report contributes to better acquaintance with and understanding of IEA educational research.

December 1983, T.J.H.M. Eggen.

Section I

Why join IEA?

J.P. Keeves
Australian Council for Educational Research
Melbourne
Australia

Chairman, distinguished guests and IEA colleagues

This year, 1983, may be considered to mark the 100th anniversary of the establishment of the field of educational research. In 1883, three events occurred which were to open up the three strands of investigation and inquiry that have characterized studies and programs of research and development in education. In that year, Stanley Hall in the United States of America published the influential book, The Study of Children, which followed the work by Preyer, a German psychologist, The Mind of the Child, which was published during the previous year in Europe. These two works marked the beginning of the Child Study Movement. Again in 1883, Sir Francis Galton published Inquiries into Human Faculty and Its Development drawing public attention to his studies on the development of tests of mental abilities. This work marked the beginning of the field of mental testing, which has laid the foundations for the Scientific Research Movement with a positivistic approach that was pursued so vigorously by E.L. Thorndike in the following decades at Teachers College, Columbia University. Also in 1883, John Dewey published the first of his major philosophical essays on 'Knowledge and the Relativity of Feeling', that was to start him on a career of philosophical study. His work, particularly that carried out at the University of Chicago, has had a profound influence on educational thought in the United States and led to the establishment of the New Education or Progressive Education Movement, in which philosophical discourse replaced the scientific approach and life experience took over from experimentation and empirical research. It is evident that these three major strands of educational research, as our colleague Gilbert de Landsheere (in press) has pointed out, the Child Study Movement, the Scientific Research Movement and the Progressive Education Movement had their beginnings in or around 1883. Consequently, it is appropriate that we, in 1983, should recognize the origins of our field of inquiry 100 years ago and pay tribute to those who inaugurated this work as well as those who have pursued their investigations so successfully in the intervening years to establish and consolidate the field of educational research.

In the period between the First and Second World Wars, a movement to establish national research institutes began and has continued during the past 50 years. Initially, institutes were set up in the sciences, particularly the applied sciences, but before long the need for work in education became evident and educational research institutes were established. The institutions founded specifically to undertake educational research include the Scottish Council for

Research in Education established in 1928, which was followed by centres set up by the Carnegie Corporation in Australia, Canada, New Zealand and South Africa. Again afer the Second World War, developed countries without such institutes established them in a variety of forms, and more recently many developing countries have seen the value of centres of this kind and have used their limited resources to create them.

It is perhaps to be expected that with greatly improved conditions for travel around the world and with technological advancements in telecommunications, a movement grew to form associations of the research centres which had been set up around the world. In education, the existence of Unesco with its three international institutes, in Paris, for educational planning, in Geneva, for dissemination of information on education, and in Hamburg, for research and scholarly work, helped to promote the idea of collaboration in educational research. Thus, it is not surprising that in 1958, exactly 25 years ago, a small group of educational research workers should, from their meetings in London and Hamburg, see the benefits to be gained from combining together to undertake research studies into common problems. As a consequence the International Association for the Evaluation of Educational Achievement was formally established a year later in 1959. During the 1960s the Association was based at the Unesco Institute for Education in Hamburg with loose affiliation to the Unesco organization in Paris. However, in the early 1970s as a direct consequence of Professor Torsten Husén's leadership and the support received from the Swedish Government, the International Institute for Education was established within the University of Stockholm and IEA, as it had become known, was housed within the informally linked to this institute.

During recent years we have seen the increased participation of educational research centres from developing countries in IEA studies and programs. However, this involvement requires considerable financial support both for the work undertaken within each participating nation as well as for travel to attend international planning and training meetings and for the work of developing a detailed research program. It would now seem possible that resources might become available through an International Fund for Educational Research in Developing Countries (IFER) to sustain within developing countries research studies that are associated with the IEA program of research in education.

The benefits of participation by developed and developing countries alike in the IEA program of research are threefold. First, there are the benefits obtained from the identification and conceptualization of a problem for research in the area under investigation. Secondly, there is the training in the conduct of research produced by instructional manuals and by following specified procedures laid down for a study, for example, in sampling and in data analysis. Thirdly, there is the important contribution that each country makes through the findings derived from the study towards an understanding of the educational process. And I would like to emphasize that we in Australia have benefited greatly in all three areas, in the identification and conceptualization of research problems, in the learning of research methods and in the building of a body of knowledge and understanding about education.

Arieh Lewy (1977) has pointed out that there are three major characteristics of IEA's research activities. First, the studies undertaken are essentially comparative in nature. The world is seen by IEA as a natural laboratory with considerable variation between countries in the conditions and circumstances within which education is conducted. Thus from the carrying out of research studies across countries it is possible to examine not only what is affecting educational outcomes within countries, but also what is influencing differences in outcomes between countries. Secondly, the studies are undertaken in a cooperative way, by educational research institutes that agree to work together to

develop a common study, to collect the basic data for the study under common conditions, and to employ common approaches in the analysis of the data and the interpretation of the findings. The sharing of findings, the frank and scholarly debate on the meaning of the findings, and the open reporting of results are a necessary consequence of the cooperative approach to research that characterizes the IEA work. It is not by the decree and direction of governments that the IEA research program proceeds, but rather by the consensus that is built up between the group of scholarly research workers from the many national centres engaged in a particular study. Thirdly, the IEA research program is firmly established within the field of empirical research in so far as it seeks generalizations that apply in one or more of the participating countries. Initially the IEA work drew upon the strategies of the Scientific Research Movement and the expertise that had been built up by Thorndike and his father at Teachers College, Columbia University. It also drew heaviliy on the methodologies and approaches to curriculum evaluation that were engendered at the University of Chicago as a consequence of the work of Tyler and Bloom and that were derived from the Eight Year Study conducted in the United States by the Progressive Education Association in the 1930s. The advent of the computer in the 1960s was both fortunate and timely for IEA, because data processing and data analysis were no longer limited by the time required for calculation by hand. Thus complex and extensive survey research, together with sophisticated approaches to causal modelling become possible under the guiding hand of Gilbert Peaker. However, the 1960s were also marked by the beginnings of an epistemological debate in educational research, perhaps in opposition to the emphasis on scientific empiricism that was being endorsed by many research workers, including those within the IEA group.

As a consequence there has developed something of a conflict between the two major paradigms that are employed in the investigation of educational problems. One is based upon the approach of the natural sciences that emphasizes empirical and quantifiable observations which can be analysed by rigorous mathematical procedures. The task of such research is to establish causal relationships and explain. The alternative paradigm is concerned with humanistic studies and is derived, in the main, from history, philoshpy and anthropology. This paradigm emphasizes qualitative information and the building of a personal interpretation of events. Clearly in the years ahead, the answer for educational research workers in IEA is not to advocate the exclusive use of one paradigm or the other, but rather to seek to employ both as appropriate.

The future of the IEA research program in all parts of the developed and developing world lies in its ability to assemble a sound body of knowledge and understanding of the educative process in order to inform and advance both educational policymaking and practice. Educational research, since the 1960s, has profited greatly from the increased resources provided for it. The IEA research program has benefited markedly from the comparative cooperative and universal nature of its activities as it has sought generalizations that will add to educational knowledge and understanding. It has been with conviction and enthusiasm that a very significant proportion of the IEA membership has contributed to the International Encyclopedia of Education, which is being prepared under the editorship of Torsten Husén and Neville Postlethwaite. This ten-volume encyclopedia is a highly significant attempt to assemble what is known about education in a coherent and readily accessible form. The preparation of the encyclopedia has not been a formal IEA activity and yet it has provided remarkable testimony of the IEA endeavour to undertake comparative research studies in a cooperative way in order to contribute both knowledge and understanding of the educative processes. We, here today as members of IEA, have joined together to do just this, and it is important for us to recognize that the publication of the International Encyclopedia of Education will mark appropriately both 100

years of educational research and 25 years of research activity by the International Association for the Evaluation of Educational Achievement.

Our answer to the question 'Why Join IEA' is that in IEA are researchers and research institutes are working together on the endless and exciting quest of searching for knowledge and an understanding of the educative process.

REFERENCES

Landsheere, G. De, History of Educational Research. In: T. Husén en N. Postlethwaite (eds). *International Encyclopedia of Education*, Oxford: Pergamon (in press).

Lewy, A., Comparative, Cooperative and Universal Aspects of the IEA Subject Matters Surveys. *International Review of Education*, 1977, 23, 305-319.

Research and policymaking in education: An international perspective.

T. Husén
University of Stockholm
Sweden

INTRODUCTION

Policy-oriented research in education covers a very short period indeed. Research delibeterately and systematically geared to provide an extended knowledge base for reform and improvement in education initiated by agents of public policy is hardly more than 25 years old. I have over the last few years had some opportunity to ponder about this in conducting a study how research and policymaking in education relate to each other in Sweden, the Federal Republic of Germany, Britain and (at the Federal level) in the United States (Husén & Kogan, Eds., in press). I have for reasons which shall not be spelled out here been able to follow what has happened in educational research as well as to study its impact on educational policy in these four countries. Therefore the study has a comparative dimension.

The comparisons have been made under two majore aspects:

1. Intra-scientific or internal conditions, such as research paradigms, schools of thought, influential researchers, and
2. Extra-scientifc or external conditions, such as availability of research funds and institutions, the "market" for research, the ideology of the state in terms of propensity for social intervention and the setting within which liaison between researchers and policymakers could be established.

I shall conlude this paper by trying to draw some lessons for the future.

The intra-scientific conditions are on the whole those which are determined by the research community itself. There were across the countries under study two overriding paradigms with dominating impact on scholarship in education: the humanistic one represented and dominated by philophers and historians, and the empirical-positivist one dominated by psychologists and - later - sociologists.

It appears convenient to distinguish two periods in the development of the disciplines that formed the basis for scholarly studies in education: the periods before and after the Second World War. When I come to extra-scientific factors, in the first place the willingness of governments to support and utilize research in education, the dividing line should perhaps be drawn at least a decade later. Policy-oriented studies in education commissioned and funded by governments began to become more frequent in the late 1950s and early 1960s. No doubt, the 1960s were the "gelden years" of educational research on both sides of the Atlantic.

INTERNAL CONDITIONS

Before 1945

In Germany the two overriding paradigms for a long time operated side by side. The philosophical, speculative approach to the study of educational problems emerged at German universities in the late 18th century when education began to be studied as a separate academic discipline with its own university chairs. The professors holding these chairs originated from philosophy. Later some had their background in history. Still around 1950 when I visited the Institute for International Educational Research in Frankfurt for a workshop about what research could do in order to improve German school education, most university professors in education, who were not many, had their background in the humanities.

Around the turn of the century empirical studies in education were conducted at several institutes of psychology. The most illustrative case is Ernst Meumann, a student of Wilhelm Wundt, who founded "experimental pedagogics" and in 1907 published "Einführung in die experimentelle Pädagogik" in three thick, impressive volumes which still were on my reading list as a young graduate student in the late 1930s. There were other leading researchers in education with their operational base in institutes of psychology, such as William Stern (1900 and 1914) in Hamburg, pioneer in educational psychology with major contributions both to differential and developmental psychology before 1914.

In the United States ever since the late 19th century, when education began to be taught at American universities, there was one predominant paradigm, the empirical one. It would suffice here to point out two or three pioneers who loomed large on the American scene. In the first place G. Stanley Hall at Johns Hopkins who, like many others, got his research training in Germany. In his "Life and Confessions of a Psychologist" he has given us a vivid picture of how educational psychology was established in the United Stages and under what paradigmatic auspices this took place. Other leading figures on the U.S. scene were Edward Lee Thorndike at Teachers College, Columbia, Lewis Terman at Stanford, and Charles Judd at the University of Chicago. The latter, who took his doctorate under Wundt in 1896, has not least in his book on "The Science of Education" made a case for education as a science in its own right, although William James already in the 1890s in his famous "Talks to Teachers on Psychology" emphatically had maintained that teaching was not a science but "an art". It appears that the low prestige that education as an academic endeavor has suffered from in the United States partly derived from the fact that the disciplinary base for educational research tended to be established outside the departments of education, whereas in many places in Europe it was established within the counterparts to these departments or in close contact with the chairs in education, some of them combined chairs in education and psychology.

The British scene before 1945 was throughout dominated by straightforward pragmatism. British universities had for a long time very few chairs of education. At Oxbridge the tradition was until recently to appoint experienced teachers and schoolmasters to these chairs because they were expected to give prospective teachers some grounding in the art of teaching.

What strikes a student of the origin of educational research in Britain is the heavy impact of the Galtonian tradition with its focus on studies of individual differences. In the laboratory in London founded by Francis Galton, at the turn of the century led by Karl Pearson, several of the leading people in the British test research were either trained or working, such as Cyril Burg and Charles Spearman. The development of intelligence tests as well as large-scale surveys by means of group tests was largely inspired by the eugenics movement that emanated from Galton (Husén, 1974). Surveys of all 11-year-olds were conducted at regular intervals in Scotland, the first one in 1933.

In Sweden there were until 1937 only three university chairs of education; a fourth was then added. Three of the incumbents were primarily experimental psychologists, thus representing the empirical paradigm. Three of them had studied with Georg Elias Müller in Göttingen, a student of Wundt. The fourth, with a background in philosophy, had studied with Bergson in Paris and Windelbank in Heidelberg, and in 1920 wrote a book on the epistemology of psychology in the Diltheyan, today one would say hermeneutic, spirit.

After 1940, when a Governmental commission of inquiry into a reform of Swedish school education was appointed, research in education gradually came in strong demand and expectations were very high both in the 1940 commission and a following one appointed in 1946 what research could do in order to provide an extended knowledge based for the reform proposals and eventually led to the introduction of the common basic comprehensive school.

The predominant influences on Swedish research in the 1940s shifted from Europe to the United States which was regarded as the Mekka not only for behavioral scientists but in particular for those wanting to absorb ideas about how to achieve progressive school reforms.

After 1945

On the German scene there was a slow re-orientation after the War with its catastrophic effects. In the 1930s many of the leading behavioral scientists had left the country, most of them for the United States. The American High Commissioner's Office made deliberate attempts to promote a change in the educational system, part of it supposed to achieve some "re-education" on the part of those, not least at the universities, with influence on the educational scene. In 1952 the High Commissioner sponsored a six-week workshop on problems of educational research at the Hochschule für internationale pädagogische Forschung which had just been established jointly by American and German authorities with the purpose of serving German schools by cross-disciplinary research in education and by long inservice training for teachers who wanted to learn research methodology appropriate for the tackling of important problems in German education. The majority of the participants in the workshop were German colleagues but there were about a dozen from other countries as well. These were expected to provide some injections from abroad. I suspect that the sponsor expected the workshop to serve as a kind of refresher course for those who had been out of contact with what had been going on outside the country in their field for quite some time.

In the early 1960s, due to the inspiring leadership and persuasive powers of Hellmut Becker (1971), a lawyer turned educator, the Max Planck Institute for Educational Research was founded in Berlin. The explicit mission of the Institute was to conduct fundamental research on a cross-disciplinary basis relevant to German problems of education that was felt to be in urgent need for reforms. There were those who at that time spoke about "twenty years of non-reform". Leading scholars at the Institute, such as Hellmut Becker and one of the pioneers of economics in education Friederich Edding, later became instrumental in the in the Bildungsrat (Federal Education Council), an organ set up to come up with recommendations for the planning of the educational system. There was in West Germany until 1970 no ministry of education and the Länder held the prerogatives with regard to educational matters. There was since the end of the 1940s the Ständige Konferenz der Kultusminister (Permanent Conference of the Ministers of Education) which was a body with itw own secretariat for mutual information and volontary cooperation.

When the constitution was changed making planning in education a Federal prerogative and when a Ministry of Education was set up, Federal support for educational research became rather abundant, at least measured by the standard of previous public support. The Länder followed suit and provided their share

of funds for both educational research and planning.

The paradigmatic pendulum among educational researchers began to swing towards a quantitatively oriented approach headed by young people with background in psychology and sociology and with training in the United States and England. A school of critical, social philosophy had been estiablshed in Frankfurt before 1933 with intellectual leadership of people such as Adorno and Horckheimer. This institute for social research assumed its activities after the war and in the 1950s a young social philosopher of the next generation, Jürgen Habermas, became the front name. The Frankfurt school played a pivotal role in the development of socialization research which has flourished at several German universities and at the Max Planck Institute in Berlin.

In the 1960s the paradigmatic pendulum began to swing back from the quantitative, positivist approach to a more humanistic and qualitative one, not least under the influence of Habermas and his colleagues. This change from measurement and quantification to understanding, hermeneutics, drew upon the humanistic-philosophical tradition of Wilhelm Dilthey, Edmund Husserl and Heidegger, the last two the leading phenomenologists. This deliberate turning the back to the neo-positivist paradigm was so fervently adopted by young German researchers that "positivist" almost became a dirty work. When some of them who came together to prepare an Enzyklopedie der Erziehungswissenschaften they seemed to have decided to make the new non-positivist, non-Anglo-Saxon approach the Leitmotiv of their encyclopedia.

It would be highly pretentious even to try to sketch what happened in educational research in the United States after the War. Suffice it to say here that in terms of paradigms the picture was pluralistic. The psychologists with their empirical approach dominated in terms of numbers, volume of research output, and recognition by the academic community. Leading scholars, such as Cronbach, Bloom, Gage and Glaser, were all trained in educational psychology. Curriculum development had slowly become a new field of study at schools and colleges of education.

In the 1950s the Federal government began to support educational research on a project basis by the Cooperative Research Program. The next injection came with the National Defense Education Act which provided big sums, not least to curriculum development, under the somewhat false label of national security. Finally, the Elementary and Secondary Act of 1965 almost over night increased the resources fo educational research manifold. Research and development centers with massive resources for tackling particular fields were set up at leading universities. Regional laboratories which were expected to be even closer to the classroom needs were established.

Given the rapidly growing support for research in education other departments than just those of education began to rally to the places were the resources were. An increasing number of psychologists were attracked to educational research as were - almost for the first time - people from other social sciences, such as sociology, political sciences and economics. Cross-disciplinary fields of inquiry were established, such as comparative education and economics of education. The diffusion of research material was revolutionized by new storage and retrieval systems, such as ERIC.

Educational researchers in Britain had in the Galtonian tradition for a long time been preoccupied by studies of individual differences, test construction, and intelligence surveys, with leading names Burt, Thomson and Vernon. Under the auspices of the 1944 Education Act and the 11+ examinations research on how to diagnose scholastic aptitude and predict school achievement became a major task for educational researchers. The social implications of the 1944 reform were in the early 1950s paid attention to by sociologists such as Jean Floud and A.H. Halsey who began to study the effects of the reform on equality of opportunity

and the extent to which parity of esteem between various secondary programs had been achieved. Since it was felt by the educational authorities that the universities did not meet the immediate needs of the practitioners, the National Foundation for Educational Research was set up as a private organization. At the beginning it was mainly a test developing institute that also conducted research on how the tests worked in schools.

In Sweden educational research from 1945 to the early 1970s was predominantly conducted by people trained in psychology. Most of their work was done in the dominant Anglo-Saxon vein with quantification and great reverence for experimental design, all according to the empirical-positivist tradition. Experimental design was the ideal, surveys second best, and observational description was regarded as a deficient substitute. But a more humanistic, hermeneutic approach, more or less closely asosciated with Marxist ideology - the so-called "rose wave" - in educational research began in the early 1970s to be propagated by a young generation of researchers.

EXTRA-SCIENTIFIC CONDITIONS

I have so far in a very sketchy way tried to convey a notion of the prevailing research tendencies and paradigms in the four countries I have studied. In what follows I shall try to identify a series of conditions outside the research community which have influenced educational research during the decades after 1945. Instead of taking country by country, which I have done in describing the paradigmatic trends, I shall take one condition at a time and in doing so compare the countries. It should also from the outset be said that there are striking similarities between the four countries in terms of how these extra-scientific factors operated. But there are also some striking dissimilarities depending upon differences in size, political system, and university traditions.

The interventionist ideology of the welfare state

Over the last few decades the state increasingly has tended to play an interventionist role in framing and implementing policies in health and education as well as welfare in general. In order to play that role successfully planning, not least in education, is necessary. In order to conduct planning an extended knowledge base is required, not only in terms of routinely collected data but also information by means of surveys, semi-experiments and analytical studies and secondary analysis of existing data.

In all the countries concerned the decades after 1945 meant a breakthrough for policy-oriented research, not least research being commisioned by governments or governmental commissions.

Prior to the early 1950s educational planning was in some places regarded as downright socialism, particularly since systematic planning had so far only been conducted in the Soviet Union. But pressure began to build up to institutionalize educational planning, particularly since international bodies, like Unesco with the establishment of the International Institute for Educational Planning in Paris and the Organisation for Economic Co-operation and Development, began to push governments to establish organs for planning inside or outside the ministries of education.

Certain educational policies that in the 1960s came to the forefront, such as provisions for better equality of educational opportunity, billingual education and education of the handicapped, could not be properly framed and implemented without information provided by surveys and evaluation studies.

Rising expectations

The "golden years" for educational research in terms of governmental support occurred in all the four countries in the 1960s and early 1970s. It was assumed that systematic and massively financed research in education would be able to do what it had achieved in industry: increase efficiency and productivity. The expectations about what could be achieved were high both on the part of researchers and policymakers. In 1971 the Select Subcommittee on Education in the U.S. House of Representatives toured Europe in order to find out what role research played in some European countries. In the Introduction to the Report from this trip the chairman of the committee, John Brademas, is quoting Charles Silberman's "Crisis in the Classroom":

"The degree of ignorance about the process of education is far greater than I had thought. Research results are far more meagre and contradictory, and progress toward the development of viable theories of learning and instruction is far slower."

Brademas points out that in defense about 10 per cent of the budget is spent on research and development, and in health 4.6 per cent.

"Yet when we come to education, as important to the life of the mind as is defense to the Nation or health to the body, we find at all levels of education in America spending an aggregate of less than one third of one per cent of their budgets on the processes of research, innovation and planned renewal.' (Educational Research in Europe, p. 3).

The Subcommittee conducted its fact-finding tour in connection with the legislation about the National Institute of Education (NIE) that was soon to be set up. NIE was thought of as a better instrument for improving American education than the system of research grants and R&D centers run by the U.S. Office of Education.

The situation by the end of the 1970s was characterized by criticism and disenchantment about education in general and about educational research in particular (Husén, 1978). This was reflected in the levelling off, or even reduction, of funds going into educational research.

Educational research conducted chiefly by social scientists was expected to provide an extended knowledge base for educational practice and policy in the same vein as did the hard science for industrial technology. What was more precisely expected varied from country to country depending upon the belief held by the elite and the general public in what science could do. In Germany, there was quite a lot of talk about "wissenschaftlichte Begleitung" (scientific accompanying) of school reforms. Even though there were academics who thought that researchers in the spirit of the Platonic philosopher-kings could come up with the full answer to how educational problems ought to be resolved, in most cases policymakers expected research to broaden their knowledge. Britain is here a particularly interesting case. Like in the other countries in the 1960s in Britain the government multipleid the resource available to social sciences with the aim of broadening the knowledge base for welfare and educational policies and their implementation. A British political scientist, Maurice Kogan, who for some time had worked in the Department of Education and Science, some years later conducted long interviews with two of the leading and most articulate ministers of education Britain ever had, Edward Boyle and Anthony Crosland. In the ensuing book, "The Politics of Education" (1983), we have the interviews on record.

Kogan characterized Edward Boyle as a "reluctant conservative", and Anthony Crosland as a "cautious revolutionary". As a formally conservative Boyle was somewhat lukewarm vis-à-vis comprehensivization, whereas "going comprehensive: was on the top of Crosland's political agenda, when in 1964 he took office as Minister of Education.

Both ministers in retrospect make reference to the research that I and my co-workers had been conducting in Sweden in connection with the school reform. Boyle deplored the short thime span available for a minister in Britain with his usually short period of tenure. He refers to Swedish Social Democratic planning which due to the stable government was able to cover "a cycle of twenty years over which a major piece of social engineering was achieved": first five years of planning, then "five years of research by Husén" (Kogan, op.cit., p. 77). Boyle evidently thought that research played a pivotal role in the Swedish school reform and regretted that given the lack of long-range political stability this was not possible in Britain. Crosland, as is clearly evidenced by Kogan's interview, also held educational research in high esteem, to the extent of inviting me to come to London in 1965 in order to meet with him for a full day when he was contemplating his famour Circular 10/65 to the Local Educational Authorities requesting plans for the re-organization of secondary education. But he held a more realistic and, in a way, more cynical conception of the role of research. In response to Kogan's question why the Circular was not preceded by research he said (Kogan, op.cit., p. 190):

"It implied that research can tell you what your objectives ought to be. But it can't. Our belief in comprehensive re-organisation was a product of fundamental value judgements about equity and equal opportunity and social division as well as about education. Research can help you to achieve your objectives, and I did in fact set going a large research project against strong opposition from all kinds of people, to assess and monitor the process of going comprehensive. But research cannot tell you whether you should go comprehensive or not - that's a basic value judgement."

But the high-strung expectations about the "answers" research was to give to basic educational problems and the ensuing improvements in educational practice were not met and therefore led to disappointment and misgivings. By the mid-1970s I happened to meet a former German Minster of Education who in Willy Brandt's government had been instrumental in increasing Federal support for educational research. During a long plane ride together the aggressively aired his misgivings about the "uselessness" of educational research. He had prior to coming into politics been a professor of mining technology and expected that the "linear" R&D model that went straightforwardly from research through development to improved mining product would work in education as well.

Educational technology

Another belief of the 1960s was the one in what educational technology based on fundamental research on the learning process would be able to do in order to make school teaching more efficient. Television, programmed learning with teaching machines and computer-based instruction in turn came on the agenda as panaceas for inefficient teaching. They all far from lived up to their promises. The fundamental reason for their failure is, of course, that education is not a manufacturing industry. In manufacuring you plan a process where you exactly know what the final products are going to be. But in education there is a wide margin of uncertainty, because its "raw material" has a wide, and largely unknown, range of potentialities. It is in the nature of the educative process of moving ahead always with a large range of options. Technology can replace teachers only to a very limited extent in that process.

In a way, the setting up of the R&D centers in the United States with support from the Federal government (Keppel, 1966) was carried by the hope that massive investment in research in a particular problem area of the kind conducted in industry would yield results that could be converted into product and methods of improved school teaching.

Funding of research

Policy-oriented research in education was practically non-existent before 1950. The following two decades saw an enormous increase of financial resources for research in education by direct government support for projects, funds available to research councils. The arrangements varied quite a lot. In Sweden, the Social Science Research Council was given considerably increased appropriations for research of a more fundamental nature. In the 1960s at the two national boards of education, the one for schools and the other for higher education, bureaus of research and development with considerable resources were established. In the Federal Republic of Germany, several central, research-promoting agencies gave support. In Britain, as mentioned earlier, most of the support went through the Social Science Research Council. In the United Strates, the Federal government took the lead.

An importang, catalytic role in bringing educational research to bear on crucial issues has in the United States been played by private foundations, such as the Carnegie Corporation and the Ford Foundation. Not only have foundations by providing initial grants to promising projects or innovations given researchers an opportunity to tackle neglected problems. They have also been instrumental in building up support for educational reforms by influencing public opinion, for instance about equality of educational opportunity. It is possible to identify several problem areas where the American foundations have taken initiatives which have subsequently been followed by support on the part of the Federal government.

In the mid-1960s the Bank of Sweden set up a research foundation, the so-called Tercentenary Foundation, of a semi-private character which also has been catalytic in supporting research in education, for instance the International Association for the Evaluation of Educational Achievement. The board of the foundation consists of six university professors and six members of Parliament.

Expanded labor market for researchers

Multiplied resources had led to a multiplication of people involved in educational research. Previously an embarrasingly large amount of research in education had been done halfheartedly by teachers who wanted to qualify for administrative positions. Given more resources young people could now invest in research careers when positions at universities within a decade doubled or trebled. This had repercussions in terms of vastly expanded graduate programs. The products of the graduate schools became employed not only at universities but at various administrative agencies as well in order to conduct surveys and other studies directly related to ongoing activities. A new category of staff policy analysts who served in a kind of liaison role between research and policymakers emerged.

Departments of education during the period under review here began to draw upon the resources offered by the whole range of social sciences. Earlier, most education departments suffered from a kind of solipcism with a focus on didactic problems and processes only. They were ready to take some help from psychology departments, but had little or no contact with other social science departments. Institutions, such as the University of Chicago and Stanford University, in the 1950s began to make joint appointment in the graduate school of education for outstanding sociologists, psychologists and political scientists. This substantially contributed to raising the quality and prestige of educational research.

Various settings for liaison between researchers and "consumers" of their products

The "consumers" of the products of educational research are in the first place practitioners in the field and policymakers in various central bodies and the administrator-bureaucrats who are expected to provide direct background material upon which decisions are supposed to be based.

Needless to say, the way research and policymaking relates in the four countries varies tremendously depending both on the size of the countries, or rather the populations, and the degree of centralization. Sweden is a special case in both respects. It has a rather small population of 8 million, and consequently the opportunities for personal contacts are much more favorable than in a country, such as the United States, with more than 200 million inhabitants. It is not too difficult for the Ministry or for the central agencies in Sweden to get together or contact most professors of education. Furthermore, the role played by the Governemnt and Parliament in Sweden in launching and promoting educational changes differs, of course, strikingly from the situation in the two federal countries, the United States and Germany. In England, local educational authorities have more influence than those in Sweden.

Liaison between research and central policymaking has been established with different models in the four countries. One is by means of blue-ribbon government commissions of inquiry, such as the so-called Royal Commissions in England and Sweden which have been highly instrumental in preparing school reforms and the legislation with them. Another model, which has been tried in a federal country like Germany, is the setting of ad hoc bodies with both academics and politicians who are expected to work out recommendations for a more uniform national policy. In the United States, White House conferences have been held focusing on important problem areas in education. Another U.S. arrangement has been the Panel of Scientific Advisors in the President's office.

In Britain and Sweden Royal Commissions constitute an important element in policy formation. Pressure begins to build up around a particular public issue, for instance better access to higher education or more equitable taxes. Representatives and advocates for the pressure groups begin to call on the Minister responsible for the particular policy area demanding that the issue should be subjected to an inquiry in dept by a Royal Commission. Simultaneously, the issue is dealt with by the media, is discussed in newspaper editorials, etc. Finally, the government gives way to the pressure or makes the judgment that it should be politically convenient to remove the issue from the forefront by "burying" it in a commission of inquiry. Such a body is usually composed of representatives of the various political parties in the Parliament, spokesmen for the organizations on the labor market and other interest groups who have a stake in the resoluation of the issue. The government gives the Commission certain terms of reference for its work, which directs its inquiry either toward a particular policy solution or leaves the field open for whatever solution the Commission might arrive at. The Commission gets a secretariat at its disposal and often conducts its own systematic fact-finding and/or research. In Sweden, for example, a considerable body of social science research over the last 30 years has been conducted under the aegis of governmental commissions.

When the Commission has submitted its main report to the government, the latter sends out the report "on remiss", for consideration and review, to various public agencies and private bodies, such as central organs of the trade unions. These reviews are submitted to the government that may decide, in case the reactions have not been too negative, to prepare legislation on the basis of the recommendations and the reactions these migh have evoked. Thus, the material provided by the Commission and the reactions by the "remiss" bodies are part of the legislative preparation. The Bill that is finally submitted for the consideration

of Parliament often quotes directly from the responses received from the reviewing bodies.

I have been going into all these details about the particular instrument of policy formation embodied in a Royal Commission in order to illustrate the very process of giving shape to a particular policy which finally is adopted by the government and the Parliament, a process also involving researchers. In times of reasonable consensus and political polarization the main issues in a particular field of policy are resolved by extended deliberations in Royal Commissions. They have to "fight it out" over a certain period of time, usually by arriving at compromises. In case the recommendations are supported only by a majority within the Commission, the minority puts its disserting opinion on record. The Secretary General of the 1957 School Commission which prepared the comprehensive school reform in Sweden was the one who prepared the draft of the Education Bill in the Ministry of Education and who, finally, served as secretary to the Parliament Select Committee that dealt with the Bill before it came up for plenary debate!

This is how the reform of the Swedish school system was prepared over a period of more than 20 years, from the mid-1940s to the late 1960s. Two Commissions, one from 1946 to 1952 and one from 1957 to 1961, dealt with the common basic 9-year school. Then, in 1960, a Commission was appointed to deal with the gymnasium, the upper secondary level covering grades 10 through 12. Alle three Commissions made extensive "use" of research (Husén, 1962; Husén, 1978). I went into more detail about "liaison" between researchers and policymakers in an AERA plenary presentation in 1965 (Husén, 1965).

In the 1960s two bodies charged with the task of providing advice in matters of reform of education and promotion of research were set up in Germany: the Bildungsrat (Education Council) and the Wissenschaftsrat (Scientific Council) (Becker in Husén and Kogan, in press). In order to ensure that they gave independent advice they were both organized according to a "two-chamber" system: one Commission of experts and one Commission with government representatives. The expert chamber of the Bildungsrat could arrive at decisions alone after having consulted the chamber with government representatives, whereas in the Scientific Council the two chambers had to arrive at decisions together. The Education Commission had a far-reaching mandate for its work. It was expected to draw up plans for German education taking into account the development trends of German society, including the manpower training demand. Furthermore, it had to make recommendations about the structure of the German school system which so far had been characterized by a high degree of parallelism according to social class. Hellmut Becker, who was vice-chairman of the Council, has from his vantage point given us an informed picture of how researchers and policymakers related in the Educational Council and its role in scientific enlightenment (Husén and Kogan, in press).

Another instrument for liaison between academics and politicians at the top executive level has been to set up Panels to advise on specific matters or on research policy in general. In 1962 the Swedish Prime Minister took the initiative in establishing what explicitely was labelled a panel of liaison between the cabinet and the research community. This body consisted of some 20-25 reputed scientists representing the whole range of research disciplines and half a dozen of those cabinet members who had ministerial responsibility for various sectors of research. The entire panel was convened once or twice a year, whereas an inner circle met more frequently to deal with more specific tasks. The secretariat was in the Prime Minister's office. Relatively little attention was, as could be expected, paid to social science and humanistic research, including research in education. This liaison body tended, however, to fade away, given the range of other tasks facing the cabinet members of the panel and the growing diversity of issues the panel had to deal with, but it has

recently been resuscitated.

In the early 1970s I had a personal experience of the Panel on Youth in the President's Office where a group of experts under James Coleman's chairmanship produced the report on "Youth - Transition to Adulthood". This group was, of course, far more removed from the concerns of top policymakers than the Swedish panel which operated in a small country with a more closely knit network of personal contacts.

CONCLUDING REMARKS AND - PERHAPS - SOME LESSONS FOR THE FUTURE

What conclusions do I draw from having been one of the actors on the scene of policy-oriented research since the 1940s and recently having devoted some studies to its role?

In the first place, there is a tendency to neglect fundamental research upon which studies of more practical problems have to draw heavily. In times of financial constraints there is a tendency to reduce the resource going to the study of more basic problems, because the damage is in the short run not as noticeable as are reductions of funds going to R&D. A proper balance has to be established between the fundamental research that forms the disciplinary basis for successful studies of a more mission-oriented or applied character.

There has been a tendency in most countries, where educational research is supported by public funds, to develop project research of an ad hoc character as a response to availability of funds for investigaging certain problem areas. By a research grant an institute or a research group within an institute can be kept alive for still a few years. The shortcomings of such a system are, of course, lack of community and narrowmindedness in the conception of problems.

There has been a growing realization of research in education primarily as an enlightnment instrument. In recent years social science researchers with their base in political science, such as Carol Weiss (1980), have studied more closely what kind of impact research has on the decision-making processes. In the first place, both her own and other empirical studies clearly show that there is not such a thing as a given piece or project of research being "fitted" into a given policy decision. Policymakers who were interviewed only in few and exceptional cases could point out how a given research had affected their stance in taking a particular decision. In the first place, decisions are seldom "taken", they emerge out of a complicated web of pressures and influences of interest groups, a process that often operates over quite some time during which no particular moment of a "decision" being made can be identified. Weiss talks about "decisions accretion". Secondly, the knowledge relevant to a particular policy issue most often derives from a multitude of pieces of research of which each contributes a tiny bit to the "knowledge creep."

Thus researchers would have to learn to play an appropriately modest role far removed from the one of pretending to act as philospher-kings. They lack both the competence and the social conditions conducive to playing such a role. At the basis of most problems of a policy-oriented nature that educational researchers are expected to deal with are value premises and political ideologies. Thus the problems cannot be solved just by presenting valid research evidence.

But researchers can in a more modest vein make important contributions to the "knowledge creep" in three respects. In the first place, they can by means of their analytical training and methodological competence be helpful in reformulating the problems and not least identify those aspects that are accessible to research efforts. They can point out aspects that tend to be overlooked. Secondly, they can serve in the enlightnment role in adding information that policymakers, practitioners, and the general public ought to pay attention to. Thirdly, and

perhaps foremost, they can serve as critics. Those of them who hold tenured positions can without being too hurt by reprisals examine the sacred cows.

REFERENCES

Aaron, H.: *Politics and the Professors: The Great Society in Perspective.* Washington, D.C., The Brookings Institution, 1978.

Becker, H.: *Bildungsforschung und Bildungsplanung.* Frankfurt, Suhrkamp, 1971.

Bildungsbericht '70: Bericht der Bundesregierung zur Bildungspolitik. Bonn, Der Bundesminister für Bildung und Wissenschaft, 1970.

Coleman, J.S. *et al.*: *Public and Private Schools.* Report prepared for the NCES, Washington, D.C., 1981.

Conant, J.B.: *Shaping Educational Policy.* New York, McGraw Hill, 1964.

Dooley, P.A. *et al.*: *Survey of Educational Researchers in Britain.* Department of Educational Enquiry, University of Aston in Birmingham, 1981.

Hall, G.S.: *Life and Confessions of a Psychologist.* New York, Appleton, 1923.

Halsey, A.H. (Ed.): *Ability and Educational Opportunity.* Paris, OECD, 1961.

Halsey, A.H.: "Social Science and Government." *Times Literary Supplement,* March 5, 1970.

House, E. *et al.*: "No Simple Answer: Critique of the Follow Through Evaluation". *Harvard Educational Review,* Vol. 48:2, 1978.

Husén, T.: "A Case-Study in Policy-Oriented Research: The Swedish School Reform." *The School Review,* Vol. 73, No. 3, 1965, 206-225.

Husén, T. (Ed.): *International Study of Achievement in Mathematics. A Comparison of Twelve Countries.* Vol. I-II. Stockholm & New York, Almqvist & Wiksell and John Wiley, 1967.

Husén, T.: *Talent, Equality and Meritocracy.* The Hague, Nijhoff, 1974.

Husén, T.: "Educational Research and Educational Reform: A Case Study of Sweden." In: Patrick Suppes (Ed.), *Impact of Research on Education: Some Case Studies.* Washington, D.C., National Academy of Education, 1978.

Husén, T.: *The School in Question: A Comparative Study of the School and its Future in Western Societies.* London & New York, Oxford University Press, 1979.

Husén, T.: "Evaluating Compensatory Education." In: *Proceedings of the National Academy of Education,* Vol. 6, Washington, D.C., National Academy of Education, 1981a.

Husén, T.: "Coleman II - Another Case of Politics and the Professors." *Change,* Vol. 13:6, September 1981, 1981b, 11-12.

Husén, T.: *An Incurable Academic: Memoirs of a Professor.* Oxford, Pergamon Press, 1983.

Husén, T & Boalt, T.: *Educational Research and Educational Change: The Case of Sweden.* Stockholm, Almqvist & Wiksell, 1968.

Husén, T. & Kogan, M. (Eds.): *Researchers and Policymaking in Education.* Oxford, Pergamon Press (in press).

Kaplan, N. *et al.*: *The Use of Social Science Knowledge in Policy Decisions at the National Level.* Ann Arbor, Michigan, University of Michigan, Institute for Social Research, 1975.

Keppel, F.: *The Necessary Revolution in American Education.* New York & London, Harper & Row, 1966.

Kogan, M.: *The Politics of Education.* London, Penguin, 1973.

Kosse, G.B. (Ed.): *Social Science Research and Public Policy Making.* London, NFER-Nelson Publishing Company, 1983.

Levin, H.: "Why Isn't Educational Research More Useful?" *Prospects,* Vol. III, No. 2, 1978, 157-166.

Meumann, E.: *Einführung in die experimentelle Pädagogik.* I-III. Leipzig, W. Engelmann, 1907.

OECD: *Reviews of National Policies for Education: France*. Paris, OECD, 1971.
OECD: *Reviews of National Policies for Education: Germany*. Paris, OECD, 1972.
OECD: *Reviews of National Policies for Education: The United States*. Paris, OECD, 1980.
Shellard, J.S. (Ed.): *Educational Research for Policy Making in Australia*. Hawthorn, Victoria, Australian Council for Educational Research, 1979.
Stern, W.: *Ueber Psychologie der Individuellen Differenzen*. Leipzig, Barth, 1900.
Stern, W.: *Psychologie der frühen Kindheit*. Leipzig, Quelle & Meyer, 1914.
Weiss, C.H.: "Knowledge Creep and Decision Accretion." *Knowledge: Creation, Diffusion, Utilization*, Vol. 1:3, 1980.
Weiss, C.H.: "Policy Research in the Context of Diffuse Decision Making." In: Kosse, G.B. (Ed.), *Social Science Research and Public Policy Making*. London, NFER-Nelson Publishing Company, 1982.

A diagnostic way of handling the Test-Curriculum overlap using Constrained Nonmetric Multidimensional Scaling.

W. de Corte
Laboratory of Metric Diagnostic and Sociopsychology of Work and Organization
State University of Ghent
Belgium

C. Brusselmans-Dehairs
Seminarie en Laboratorium voor Didactiek
State University of Ghent
Belgium

INTRODUCTION

International comparative studies on program evaluation encounter a major problem while attempting to relate the multiplicity of curricula to a single, limited test battery. Direct comparison of the national achievement scores is seriously hampered by the differences between nations with respect to what is tested and what is taught. This problem was recognized from the beginning by the International Association for the Evaluation of Educational Achievement (IEA). In its first mathematics survey (Husén, 1967), a measure of opportunity to learn (OTL) was devised to capture the extent to which the intended curriculum, situated at the Educational System level of focus, had been implemented in the individual classroom (the implemented curriculum, situated at the Classroom level of focus (Travers, 1980)). Further refinements of the measure, inspired by the objections raised among others by Freudenthal (1975), led to a somewhat different content coverage. The revised instrument, called "opportunity to learn content of the IEA test" (Travers, 1980, p. 196) now focuses on how much of the subject matter of the test is taught by the teacher. In this sense the new measure essentially coincides with the notion of overlap, as discussed by e.g., Leinhardt and Seewald (1981, p.85): "By overlap, we mean the extent to which there is a match between the content of what is taught and the content of the test used to measure progress in performance."

For the above stated problem, being recognized for quite some time (Cole & Nitko, 1979; Comber & Keeves, 1973; Husén, 1967; Rosenshine, 1978; Walker, 1976), mainly two approaches to the test-curriculum overlap problem have been worked out. The curriculum based strategies try to eliminate the lack of fit between curricula and tests either by developing new "criterion based" tests (Hambleton, 1982; Leinhardt & Seewald, 1981; Popham, 1979), by altering existing tests, or by equating curriculum and test content on the basis of taxonomic analysis (Armbuster, Stevens & Rosenshine, 1977; Floden, Porter, Schmidt & Freeman, 1978; Kuhs, Schmidt, Porter, Floden, Freeman & Schwille, 1979; Steiner, 1980). The measurement approach takes the overlap for granted, but incorporates into the achievement analysis some numerical estimate of the relationship (e.g., Cooley, Leinhardt & Zigmond, 1979; Konttinen, 1981; Leinhardt & Seewald, 1981).

Although both strategies exhibit important intrinsic and pragmatic differences neither of them is really up to the specific problems posed by international comperative research. The elimination, or adjustement approach can only partially be adopted when multiple divergent criteria are present. Although the fit between the curricula and the tests can be maximized by careful taxonomic analysis, important differences in the extend of content coverage remain (see tables 12.1 and 12.3 of the IEA bulletin No.4, 1979). An additional direct measurement of these differences, i.e., a measurement approach is required. However, the current ways in which these scores are subsequently used - e.g., to explain part of the between country differences - do not permit the type of qualitative insight gained by the elimination methods. To overcome some of these relevance problems, we will expose a new methodology for relating the OTL information to the achievement scores. The methodology tries to combine the advantages of the qualitative, curriculum based approaches with the straightforward but rather uninformative measurement solutions. A brief survey of the latter practices will provide the rational for our method.

MEASUREMENT APPROACHES AND COMPARATIVE RESEARCH

A first practice aggregates the OTL scores, obtained at the item level, to construct a global OTL index for each pupil. This index is then correlated with the achievement score (Comber & Keeves, 1973; Konttinen, 1981; Leinhardt & Seewald, 1981; Walker, 1976). Aside from the presuppositions made - i.e., that all the potential topics and tasks are covered and well sampled by the items (Konttinen, 1981, p. 2) - all that results from the operations in the best of cases is the confirmation of a highly expected phenomenon; exemplifying by this the kind of trivial and uninformative conclusion we mentioned above.

A second measurement approach aggregates the OTL indices and item scores over students, and studies their covariation over items (Konttinen, 1981). Again a positive relationship is expected; and again its empirical realization is quite trivial a finding. Less trivial is the finding of low, but, due to the large number of cases, eventually significant correlations (e.g., Konttinen, 1981), suggesting that either the specific numerical treatment of the OTL scores, the implicit rational underlying the analysis, or the operationalization of the OTL criterion is inappropriate.

Still other ways of dealing with the OTL or overlap measures have been conceived. Leinhardt and Engel (1982); Leinhardt and Seewald (1981) and Leinhardt, Zigmond and Cooley (1982) used a regression approach. In the regression equations overlap, besides pretest and process information (e.g., Instruction time, teacher behaviors, etc.), is used to predict the posttest scores. Konttinen (1981) performed multiple classification analysis relating item OTL type to item difficulty. The OTL types (item content given this year ; T; given earlier : E; later : L; etc.) referred to the number of teachers that gave the rating to the item. Although a multiple R of 0.50 is obtained, Konttinen (1981, p. 8) concludes that : "... even in combination the item OTL measures have only weak relationships with item difficulty." The same author also applied logit analysis of variance, i.e. linear regression of the correct responses on OTL with logit link functions and a binomial error model. In only 11 of the 176 analysed cases did the model fit the data.

In summary, the results of the measuring approach are rather disappointing. Above we gave a number of hints as to why this could be the case. More specifically, we do indeed believe that the OTL data contain relevant information with regard to the achievement responses. But to obtain substantial data

on the nature of this information, we should not focus on the relationship between the OTL and achievement scores, both perceived as unidimensional constructs. Instead we should concentrate on the structural analogy of the OTL and item response space; and we should do this with due respect to the measurement level of the structural indices we can reasonable expect to obtain. Indeed most of the items, judged by the teachers, and responded to by the pupils, are of a complex nature, and are based on several constituent bits of knowledge. The items may e.g., differ both in content, and degree of importance. They generally also refer to different behavioral categories. So the global ratings of the teachers may very well express a multi-facetted judgement, whatever the formulation of the OTL question. The answer of the pupils may in the same way express a multidimensional phenomenon. From a diagnostic, learn-theoretic point of view, it would be all the more interesting to have an idea of the underlying aspects structuring both the OTL and the performance space. This transforms the question about the relationship between the OTL and the test scores to a question pertaining to the structural similarity between the OTL and performance universes.

THE STRUCTURAL SIMILARITY BETWEEN THE OTL AND THE PERFORMANCE SPACE

To solve the structure similarity problem, we propose the following three step procedure. First we specify how to obtain the raw structure data on both OTL and achievement. We then indicate how information about the underlying aspects of the OTL and the performance space can be extracted. Lastly we propose a method to check the structural similarity. In all three steps we pay special attention to the measurement level of the data at hand.

A very simple and popular way of obtaining the raw structure data consists of constructing two similarity data matrices : one with regard to the OTL responses; the other referring to the achievement scores. For each item couple we calculate a similarity coefficient; whereby the type of coefficient used is dependent upon the nature of the OTL and the achievement (ACH) measure. For interval data we propose Pearson's product moment correlation. For ordinal data the polychoric correlation (Olson, 1978), or the Goodman-Kruskal gamma is suited (Napior, 1972); and for nominal data we have the choice between a number of association measures. In all cases the similarity coefficients are computed over teachers (OTL), and over pupils (ACH).

In the next step we attempt to find the dimensions that underly the OTL (or the ACH) structure data. Here again we can chose among a number of alternatives. In the application of the method, discussed underneath, we use non-metric multidimensional scaling (MDS). Another option is factor analysis. Although both methods have been criticized in the past on the grounds that their explorative applications seldom resulted in a substantial contribution to the field of interest (e.g., Shepard, 1974), their use is very well established. Furthermore, we outline in the discussion section a non-explorative approach, which combines steps 2 and 3 of the present method, and which is not longer subject to the invariance problems associated with classical factor analysis and MDS solutions.

The last, and crucial phase essentially consists of verifying whether the structure found with regard to the OTL (or the ACH) data is also exemplified by the ACH (or the OTL) similarity measures. Again we propose the use of MDS. But instead of performing an explorative search, we will use a recently developed extension : constrained MDS (De Corte, 1982). Unlike the popular MDS techniques, constrained MDS (CMDS) produces a scaling solution which is not only based on the similarity data; but which also takes into account some

prespecified hypothesis with regard to the underlying dimensions. The hypothesis more specifically refers to the picture - i.e., the extracted dimensions, and the order of the scaled items on each of these axes - obtained at stage two of the global procedure.

Essentially two circumstance favor the adoption of CMDS. The first circumstances which is negatively inspired is related to some of the current problems associated with ordinary MDS. These problems have been discussed at substantial length by e.g., Shepard (1974); Borg (1981); and De Corte (1982). Most relevant in this context are the difficulties with regard to the local minima, the general indeterminacy of the solution configuration, and the resulting interpretation problems. In fact, the negative circumstance is closely related to the experiences which shadowed the use of explorative factor analysis.

The second, positively oriented circumstance has to do with the very common observation that we generally dispose of more information than merely the similarity data to obtain a scaling solution. This additional information may take on a number of forms; thereby specifying the type of CDMS that will be needed. Different types of CMDS have among others been proposed by Noma and Johnson (1979) and Borg and Lingoes (1980), while de Leeuw and Heiser (1980) discuss a very general algorithmic scheme.

We favour the use of additional ordinal restrictions on the point coordinates (i.e., the dimensions). As is explained elsewhere (De Corte, 1982), we believe that the adoption of a distance model for analyzing similarity data (as is the case in ordinary MDS and in CMDS) is not entirely at par with a regional (cluster) or manifold (simplex, circumplex, tec.) oriented interpretation mode. Although the dimensions arrived at by exploratory MDS are arbitrary, a dimensional interpretation still is the natural way of looking at an appropriately constrained MDS solution. The choice for ordinal restrictions on these dimensions in its turn is based on the fact that the level of theory building in the educational sciences is not that sophisticated as to warrant the use of more fine-grained, linear restrictions. It e.g., suffices to look at some of the leading taxonomic dimensions (Bloom, 1956; De Block, 1975) to find that, although the order between the levels of a certain taxonomic aspect (e.g., behavioral category) is clearly specified, no indication whatsoever is given as to the intervals spacing the instances. The eventuality of categorical taxonomic aspects will be discussed in a subsequent paragraph.

In summary, we propose the use of both constrained and unconstrained MDS to solve the problem of assessing the structural similarity between the OTL and ACH data spaces. The global procedure retains a good deal of the qualitative approach in that learn-theoretic, and taxonomic principles, governing the data, can reveal themselves. It is precisely at this level that the relationship between OTL and ACH is investigated, and not, as is usual in the ordinary measurement approach, at the (aggregated) raw data level. Finally, the relationship is analyzed by means of a method - i.e., MDS with order constraints on the point coordinates - that does not violate the somewhat crude standards of the current educational and didactical theories.

MDS WITH ORDER CONSTRAINTS ON THE POINT COORDINATES

The technical machinery of our CMDS method has been presented by Noma and Johnson (1979), Borg and Lingoes (1980), and De Corte (1982). Instead of a recapitulation, we will focus on the concrete implementation of the method for the purposes at hand. Suppose one is interested in whether or not the ACH similarity data reveal the same qualitative structure as the one that shows up in an ordinary MDS of the OTL scores. One way of translating the latter

qualitative structure is by enumerating the order of the items on each of the axes in the obtained solution. We consider the ACH data to have the same structure as the OTL data whenever they can be scaled according to the above specified constraints, such that the fit between the original similarity data and the scaled distances, is not markedly less than under the unconstrained scaling condition.

The criterion, specified for the notion of structure similarity, might be too stringent in a number of applications. It may sometimes be reasonably to expect only a partial resemblance between the OTL and the ACH structure; the partiality either occuring at the dimension level, at the global solution level, or pertaining to a combination of both. In a real application a total of three dimensions may be required to display the OTL data. The first dimension reflecting the importance, the second the behavioral category, while the third dimension can not be meaningfully interpreted within a learn-theoretic or didactic perspective. In such a case it would be acceptable not to impose constraints on all three dimensions. Using only the restrictions on dimensions 1 and 2, we then obtain an example of partiality at the global solution level. When we furthermore regroup the obtained order of the n items on the first dimension into a smaller set of ordered categories, and eventually delete several items from the new order - the importance dimension in that way being more neatly exemplified - we have an instance of partiality at the dimension level.

The above example made no doubt clear that the notions of partiality far more reflect a qualitative necessity than a technical possibility. Full-blown application of CMDS for the qualitative oriented investigation of the test - curriculum overlap cannot materialize without the scaling technique being up to these kind of situations. Suffice it to note that the algorithm performing the constrained scaling - i.e., the CDM technique (De Corte, 1982) can indeed handle any of the forementioned instances.

Before turning to a pilot application of the three stage procedure, we discuss a possible alternative procedure. It could be argued that there is no real necessity for CDMS in the third phase : instead ordinary MDS could be employed. A fourth stage would then be introduced to check whether the dimensions, found in step 2 can be fitted in the picture of stage 3. Property fitting (Chang & Carroll, 1970; Kruskal & Wish, 1978) is one such technique. Without going too deep into the matter, the essential problem with such an approach is a) that the solution configuration depends on the initial configuration, and, related to the former issue, that b) the solution often is a local minimum. Several authors have shown (e.g., Noma & Jognson, 1979; Borg, 1981; De Corte, 1982) that for a given data matrix qualitatively different solutions - i.e. solutions which can not be related to each other by means of the set of acceptable transformations - may be obtained; all having approximately the same fit value (i.e., they pertain to different but equivalent local minima). Suppose that one or more of these local minima is associated with solutions that indeed reflect the structural similarity, while the others do not. In that case there is a substantial (but inestimable) chance for the similarity not to be revealed, when the algorithmic search process does not take into account the external information concerning the structural hypothesis. An appropriate CMDS technique, on the other hand, will, under these circumstance, guide the solution to one of the local minima within, on, or nearby (depending on the way the restrictions are implemented) the feasible region - i.e., the part of the solution space where the restrictions are met. Moreover the fit of the CMDS solution will not be markedly different from the one associated with an unconstrained representation. In other words, CMDS implies a substantial reduction in the chance of a faulty rejection of the structure similarity hypothesis, compared to the alternative approach.

A PILOT APPLICATION OF CONSTRAINED MDS

Data.

The data used come from the Belgian (Fl.) population B sample (i.e., students from the last year of higher secondary education, that have at least five hours of mathematics a week), that was studied as part of the second IEA mathematics study. We will not comment at substantial length on the sample. For a description see Brusselman-Dehairs (1981). Suffice it to note that some 600 pupils, proportionally distributed over the different national types of population B education, responded to the ACH test.

The raw OTL measures, obtained at the item level, refer to the estimates by the national education inspectors of the percentage of population B students that are expected to pass the items. The estimates were given on the basis of whether or not the item content, the behavioral categories implied by the item, etc. appeared in the intended curriculum.

The items, 22 in total, that were investigated, constitute one of the parallel forms worked out in the IEA population B design. We took a form which did not only contain questions from the international core, but which was also augmented with five national option items. Apart from the binary pupil by item matrix of individual achievement responses, we furthermore disposed of the difficulty indices, calculated on the basis of the national sample.

Analysis.

The structure similarity hypothesis, concerning the OTL and the ACH data spaces, was checked by means of the three step procedure explained above. After the construction of both the OTL and the ACH similarity data (we used Goodman-Kruskal's gamma for OTL and tetrachoric correlation for the ACH data), the MINISSA algorithm (Roskam & Lingoes, 1970) was employed to reveal the underlying aspects of the OTL similarity coefficients. We found that an adequate representation could be given in a three dimensional space. The stress (S = .046), associated with this three dimensional solution (3D solution), indicated that the representation is excellent (Kruskal & Wish, 1978).

The unconstrained 3D representation of the OTL structure data was then converted into a set of order constraints on the point coordinates; thereby specifying the first external hypothesis (H1) with regard tot the ACH similarity data. In a second approach (H2), we added a fourth dimension specification, pertaining to the difficulty level of the items. Table 1 illustrates the translation process to obtain the order restrictions for both the H1 and H2 scaling. Each first row of the first three blocks of table 1 refers to the MINISSA 3D coordinates for the OTL data. The second row within each block summarizes the corresponding ordinal translation. The rank numbers indicate the order that will be implemented on the ACH items for each of the restricted dimensions. The fourth block illustrates the conversion for the difficulty continuum.

Hypothesis	Stress	Correlation (R) data - solution	Correlation (ξ) restrict dim. coordin.	Borg - Lingoes dj - test ($\bar{df}$ = 228)
H0	0.130	0.862	n.a.	n.a.
H1	0.178	0.782	0.725 0.673 0.745	2.0325 $p > 0.01$
H2	0.193	0.719	0.719 0.787 0.654 0.764	3.5402 $p < 0.01$

Table 2. Comparison of the accuracy with which H1 and H2 describe the structure underlying the ACH data.

between the ACH similarity values (the data), and the corresponding MDS inter-point distances, after both were transformed to rank numbers. The index specifies the goodness of fit of the representation with regard to the data. The quantities in column 4 are rank correlations (Kendall tau), which indicate for each of the constrained dimensions the goodness of fit with the ordinal restrictions.

The results in Table 2 show that while H2 is too stringent, the ordinal restrictions, related to the OTL dimensions (H1), provide an acceptable basis for structuring the ACH data space. Accordingly, the conjecture about the structural analogy between the OTL and the ACH similarity data is confirmed.

Although it is not of central importance to the relevance of the methodology put forward in this paper, we tentatively labeled one of the three order restrictions which constitute H1. The first order, corresponding to the first MDS solution dimension of the OTL data, could be traced back to the taxonomic work of De Block (1975). While the more widely known taxonomy of Bloom (1956) is adopted in the IEA study to structure the cognitive items with regard to "behavioral level", it is a fact that the classification of De Block is much more repanded in the Flemish part of the Belgian country.

As a final illustration of the obtained results, we present in Figure 1 the constrained MDS solution of the ACH data, superposed on the unconstrained representation of the OTL similarity data. Only dimensions one to three of the ACH space - i.e., the externally constrained aspects - are represented.

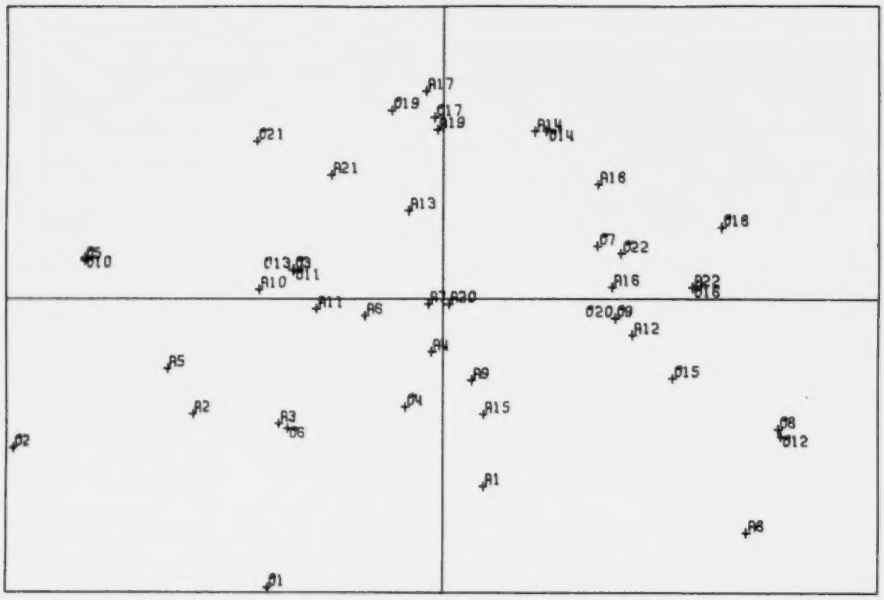


Figure 1a. Joint representation scaling OTL data and constrained scaling (H1). ACH data - dimensions 1 and 2.

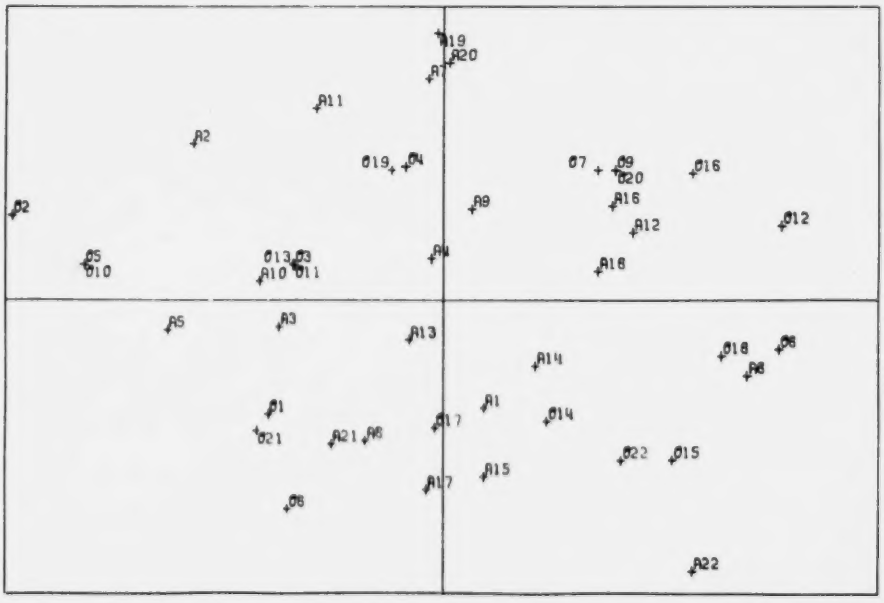


Figure 1b. Joint representation scaling OTL data and constrained scaling (H1). ACH data - dimensions 1 and 3.

DISCUSSION

Our pilot application closely followed the three step procedure outlined above. At that time, we mentioned the possibility of a completely non-exploratory approach. We will first comment on that issue. A brief note on what to do when the external hypothesis can not be translated into order restrictions, will conclude the paper.

The main drawback of the present proposal is associated with the indeterminacy of the unconstrained representation obtained at step two. Nothing guarantees that the classical scaling of the OTL data space will result in the exemplification of theoretically interesting aspects. However, as indicated above, the use of constrained MDS need not be restricted to the third phase. If the researcher has some definite guesses as to which important principles govern the interrelationships whithin the OTL and the ACH space, he could just try constrained scaling of both data sets. He then has two comparisons to make. The first comparison relates to the goodness of fit with regard to the OTL data. The second relates to the ACH data. Each comparison necessitates a) the construction of a baseline - i.e., performing an unconstrained scaling in a space of appropriate dimensionality -, and b) the evaluation of the constrained solution, using e.g., Borg & Lingoes' t-statistic. Whenever both comparisons turn out favourably, the researcher is provided with the confirmation of a qualitative, and generally theoretically and pragmatically important insight on the nature of the OTL-ACH relationship.

What to do when certain facets of the external hypothesis withstand the translation into ordinal restrictions? Bahavioral level, degree of appropriateness, centrality, etc. are all characteristics which lend themselves quite naturally to an ordinal specification. This is however not the case with e.g., domain of interest, or content category. The latter principles clearly refer to some sort of partitioning for which an ordering of the contained equivalence classes makes very little or no sense at all. Although we do not advocate the use of a distance model, and consequently, the implementation of a MDS algorithm, for these cases, it might under certain circumstances nevertheless be practical to perform a constrained scaling. Especially when the external hypothesis is of a mixed format: some restrictions referring to an order relation; some others pertaining to a partitioning. The algorithm, used for the pilot application was extended to cope with this kind of situation.

In summary, we believe that some of the recent developments in the area of MDS, can indeed be integrated in order to close the present gap between tests (the criterion) and curricula. From a pragmatic point of view, the investigation of test-curriculum overlaps should result in pinpointing those which not only are effective, but which are also alterable. Constrained MDS precisely helps in the confirmatory search for such aspects.

REFERENCES

Armbuster, B.B., Stevens, R.J. & Rosenshine, B.: *Analyzing content coverage and emphasis: A study of three curricula and two tests* (Technical Report 26). Urbana, IL: University of Illinois at Urbana - Champaign, 1977.

Bloom, B.S.: *Taxonomy of educational objectives*. New York: David Mc Kay Comp. Inc., 1956.

Borg, I.: *Anwendungsorientierte Multidimensionale Skalierung*. Berlin: Springer Verlag, 1981.

Borg, I. & Lingoes, J.C.: A model and algorithm for multidimensional scaling with external constraints on the distances. *Psychometrika*, 1980, *45*, 25-38.

Borg, I. & Lingoes, J.C.: An alternative approach to confirmatory inference and geometric models. *Quality and Quantity*, 1981, *15*, 541-552.

Brusselmans-Dehairs, C.: *Sampling plan IEA-Second Mathematics Study, Belgium (Flemish)*, State University of Ghent, 1981.

Chang, J.J. & Carroll, J.D.: *How to use PROFIT, a computer program for property fitting by optimizing nonlinear or linear correlation.* Murray Hill, NJ: Bell Labs, 1970.

Cole, N.S. & Nitko, A.J.: *Instrumentation and bias: Issues in selecting measures for educational evaluation.* Paper presented at the National Symposium on Educational Research, Johns Hopkins University, November 1979.

Comber, L.C. & Keeves, J.P.: *Science education in nineteen countries. International Studies in Evaluation I.* New York: John Wiley & Sons, 1973.

Cooley, W.W., Leinhardt, G. & Zigmond, N.: *Explaining reading performance of learning disabled students.* Pittsburgh, PA: University of Pittsburgh, Learning Research and Development Center, 1979. (LRDC Publication No 1979/12).

De Block, A.: *Taxonomie van leerdoelen.* Antwerpen: Standaard Wetenschappelijke Uitgeverij, 1975.

De Block, A. & Brusselmans-Dehairs, C.: *Nationale ondersoeksinstrumenten studie wiskunde in het kader van de Belgische deelname aan het ondersoek van de Internationale Vereniging voor de Evaluatie van het Schoolrendement (I.E.A.)* - Transversaal project populatie B - Ministerie van Nationale Opvoeding en Cultuur, 1981.

De Corte, W.: *Restrictief multidimensioneel schalen. Een bijdrage tot het variabelen-taal onderzoek in de A&O psychologie.* Unpublished doctoral dissertation, State University Ghent, 1982.

De Leeuw, J. & Heiser, W.: Multidimensional scaling with restrictions on the configuration. In P.R. Krishnaiah (Ed.), *Multivariate Analysis.* Vol. V. Amsterdam/New York: North Holland, 1980.

Floden, R.E., Porter, A.C., Schmidt, W.H. & Freeman, D.J.: *Don't they all measure the same thing? Consequences of selecting standardized tests.* East Lansing, MI: Institute for Research on Teaching, Michigan State University, July 1978. (Research Series no. 25)

Freudenthal, H.: Pupils achievements internationally compared - the IEA. *Educational Studies in Mathematics, 6*, 1975, 127-186.

Hambleton, R.K. (Ed.): Contributions to criterion-referenced testing technology. *Applied Psychological Measurement*, 1982.

Husén, T. (Ed.): *International study of achievement in mathematics: A comparison of twelve countries.* Volumes I and II. New York: Wiley & Sons, 1967.

International Association for the Evaluation of Educational Achievement. *Second study of mathematics*, Bulletin 4, december, 1979.

Konttinen, R.: *Some reflections on the opportunity - to - learn and score equating.* Jyväskylä, Finland: Institute for Educational Research, University of Jyväskylä, 1981. (Memorandum)

Kruskal, J.B. & Wish, M.: *Multidimensional scaling.* Beverly Hills/London: Sage, 1978.

Kuhs, T., Schmidt, W., Porter, A., Floden, R., Freeman, D. & Schwille, J.: *A taxonomy for classifying elementary school mathematics content.* East Lansing, MI: Institute for Research on Teaching, Michigan State University, April 1979. (Research Series No. 4)

Leinhardt, G. & Engel, M.: *An iterative evaluation of NRS: Ripples in a pond.* Paper presented at the Annual Meeting of the American Educational Research Association, Boston, 1980.

Leinhardt, G. & Seewald, A.M.: Overlap: What's tested, what's taught? *Journal of Educational Measurement*, 1981, *18*, 85-96.

Noma, E. & Johnson, J.: *Constrained nonmetric multidimensional scaling*. Ann Arbor, MI: University of Michigan Math. Psych. Program, 1979.

Popham, W.J.: The case for criterion-referenced measurements. *Educational Researcher*, 1978, *7*, 6-10.

Olson, V.: Maximum likehood estimation of the polychoric correlation coefficient. *Psychometrika*, 1978, *44*, 443-460.

Rosenshine, B.: *Academic engaged minutes, content covered, and direct instruction*. Unpublished manuscript, University of Illinois at Urbana - Champaign, 1978.

Roskam, E.E.: *Measurement of attitudes*. Intensive advanced course paper. Brussels: EIASM, 1982.

Roskam, E.E. & Lingoes, J.C.: MINISSA-I: A Fortran IV (G) program for the smallest space analysis of square symmetric matrices. *Behavorial Sciences*, 1970, *15*, 204-205.

Shepard, R.N.: Representation of structure in similarity data: Problems and prospects. *Psychometrika*, 1974, *39*, 373-421.

Steiner, H.G. (Ed.): *Comparative studies of mathematics curricula - Change and stability 1960-1980*. Institut für Didaktik der Mathematik der Universität Bielefeld. Materialen und Studien, Band 19, 1980.

Travers, K.T.: The second international mathematics study: an overview. In Steiner, H.G.(Ed.): *Comparative studies of mathematics curricula - Change and stability 1960-1980*. Institut für Didaktik der Mathematik der Universität Bielefeld. Materialen und Studien, Band 19, 1980.

Van Den Ende, H.W.: *Beschrijvende statistiek voor gedragswetenschappen*. Amsterdam/Brussel: Agon Elsevier, 1971.

Walker, D.A.: *The IEA six subject survey: An empirical study of education in twenty-one countries. International Studies in Evaluation, Vol.9*. Stockholm: Almqvist & Wiksell and New York: John Wiley, 1976.

Section II

The Second International Mathematics Study.

General Overview

R.W. Phillipps
Department of Education
Wellington
New-Zealand.

It is with some trepidation that I approach the task of giving, within a five minute presentation, an overview of a research project that has fully occupied the attention of very many researchers around the world for the last seven or eight years.

From the outset it is important to recognise that this study has attempted to address the issues facing mathematics educators rather than attempting to use mathematics as a surrogate for a more general measure of national achievement. The fundamental question behind the study is to ask that if there are national or international differences in mathematics achievement wat might it be within the system or the classroom that contributes to these differences. The study has therefore concentrated its attention on three elements. It has attempted to gather information which will enable an accurate desciption to be made of each country's official curriculum, ie. the Intended Curriculum. As well as gathering a wealth of information of a descriptive nature the study has attempted, through the developement of international grids, to achieve a consensus judgement of the importance accorded topics in mathematics by each country and hence to codify each country's curriculum. Even though this is a somewhat crude instrument it does appear that clusters of countries can be identified which have common curricular influences and history. Work on this aspect of the study is continuing in Urbana as part of the Curriculum Analysis Report. These international grids also formed the framework for the selection of the items in the cognitive test. To assess the match of the tests to each country's intended curriculum, the National Mathematics Committees were asked to rate the appropriateness of each item for their country - in other words were the items acceptable. Clearly this measure needs to be considered whenever an attempt is made to assess the cognitive results if the tests are to be kept in the correct perspective.

The second element was the attempt to indentify the Implemented Curriculum as it occurs in the classroom and to study its relationship to the Intended Curriculum. In other words do the official syllabus statements reflect what is actually taking place in the classroom. One of the main measures intended to assess the implemented Curriculum is the measure of opportunity to learn and a discussion of this measure as it affects The Netherlands is the basis of the paper being presented this afternoon.

The third element of the design is the Achieved Curriculum as portrayed by the student outcomes - both cognitive and affective. The international Specialists Committee responsible for the construction of the cognitive measures was very conscious of previous criticisms of the first mathematics study instruments.

In an attempt to provide as wide a curriculum coverage as possible in this study, while at the same time keeping the testing time to a minimum, item sampling was used. This approach does pose problems for the creation of individual student scores and hence any traditional IEA between-student analyses. However, the approach will give stable measures for national and school or class means.

The outcome measures are supported by extensive school, teacher and student questionnaires with a number of attitudional measures. Some countries also administered a measure of the students' perception of whether they had had an opportunity to learn the mathematics behind each item as well as an indication of whether or not they used a hand calculator to solve an item.

And student as mentioned earlier the specialist committee was particularly interested in what went on within the classroom and the variables which could contribute to the explanation of any variance in the outcomes. Initially, it was hoped that all countries would, at Population A, conduct a longitudinal study with a pretest and posttest as well as the administration of extensive questionnaires to the teachers about how they approached the teaching of 5 mathematics topics which seemed to be of interest in most of the countries. A number of factors including the funding commitments of individual countries and the doubts they held about the willingness of their teachers to cooperate with such an extensive set of questionnaires has meant that only eight countries eventally took part in the longitudinal aspect of the study. The initial analyses of these data are showing promising leads in the identification of growth patterns and their explanatory variables.

As can be appreciated this study has generated a mammoth amount of data Within the very limited resources which have been available to the study every effort is being made to archive to data in a bank supported by fully documented record of the status of each piece of data. It is hoped to make this bank available around the world for secondary analyses.

The first results of the study are to be published in a series of reports The first three official reports cover:

a. the Curriculum Analyses
b. the results from the Cross-sectional Study and
c. the Classroom Processes from the Longitudinal Study

And it is hoped that these will be available over the next two years.

Some results of the second international mathematics study in The Netherlands.

Th.J.H.M. Eggen, W.J. Pelgrum, Tj. Plomp
Twente University of Technology
Department of Education
The Netherlands.

INTRODUCTION

From November 1977 (till March 1983) the Netherlands was participating in the Second International Mathematics Study (SIMS) of IEA. In this paper only a global description of this project and its results are presented. For more detailed information we refer to the project proposal and other project publications (see references). One of the possible uses of the data gathered in the study is illustrated in this paper: viz. the description of several aspects of mathematic curricula.

The SIMS is an IEA project. IEA is an international organization with about 40 member countries. Since the early sixties IEA has been involved in multinational research projects. At first, attention focussed on the study of the outcomes of the education in several disciplines. In recent projects a wider range of educational research questions such as the causes of early school leaving and the developement of an international item bank has been studied. Twelve countries took part in IEA's first project: the first mathematics project. The results of this study are reported internationally by Husén,(1967). Wiegersma and Groen (1968) reported the results of the Dutch participation.

In the period 1970-1975 the Six Subject Study was undertaken. This investigated reading comprehension, science, civils, English (as a foreingn language), French (as a foreign language) and literature. The results of this study are are reported in the 9 volumes of the International Studies in Evaluation, while the Dutch results on science and mother tongue are reported by Sandbergen (1974).

BACKGROUND OF THE SECOND MATHEMATICS STUDY.

In the sixties important changes in the mathematics educations took place all over the world. Changing opinions about the content and the didactics of school mathematics were the starting point of a profound revision of the mathematics curricula (see e.g. Treffers, 1978, for a description). In many countries these developements stabilized in the beginning of the seventies. The second part of this decade is therefore a good period for a state-of-the-art study of mathematics in the schools.

The major aim of the project is to give a description of the relationships which exist between

a. The mathematics program (what is the content and the context of mathematics teaching?),

b. The affective and cognitive results of the students (what is the output of mathematics teaching?) and
c. The teaching-learning process (in what way is the output achieved?).

We can study the mathematics curriculum on three different levels. On the first level we have the intended curriculum, as specified in the official documents of a country. The second level is the curriculum as implemented within the schools and the classrooms. In the actual mathematics lessons the intended curriculum is given its concrete form. Here the time to be spent on the parts of the curriculum, the didactics and the methods are determined. Finally, we have the attained curriculum: the (affective and cognitive) objectives the students have attained. In the study the content of each of these levels is described and the relationships between them are investigated. Each curriculum level is a special object of study in certain parts of the SIMS (see fig.1). In this figure is also indicated on which level data are collected.

	Study component	Object of Study	Data from
I	Curriculum-analyses	Intended Curriculum	Countries (educational systems)
II	Classroom processes	Implemented Curriculum	School and Class
III	Outcomes	Attainded Curriculum	Student

Figure 1 : Schematic view of the study.

In the curriculum analysis part of the study, attentions is paid to the content (i.e. the topics on school mathematics) and the context (e.g. school system, examination system) of the intended mathematics curriculum. In this paper we will not deal with these analysis; see Steiner (1980) for the first results.

The study of the teaching-learning processes within the classrooms is (amongst others) directed tot the desciption of the implemented curriculum, the methods used and the didactics applied in this methods.
In the third part of the study the cognitive and the affective results of the students are assessed in relation to the intended and implemented curriculum and several other variables (e.g. hours spend on home work and gender).

SUMMARY DESIGN AND INSTRUMENTS.

In the next sections only those data about the design of the study are mentioned which are necessary for a good understanding of the results presented later.

The design of the study.

21 countries participated in the SIMS. The design of the study was a result of discussions between the participating countries. Each country could take part according to the complete international design or only in parts of the study.

The Netherlands decided for a limited participation in the SIMS, by restricting itself to one of the two internationally proposed populations. The international definition of this population (population A) is: all students in the grade level where the majority has attained the age of 13.00-13.11 by the middle of the school year. In the Netherlands this population was

determined as the second year of secondary education (US-grade level 8). In the Dutch school system a number of different schooltypes can be distinguished at this grade level. First of all we can distinguish between school types which offer a general education and school types which offer elementary vocational education (LBO).

Schooltype	Enrolment % in level 8 (N=276.807)
Pre-university education (VWO)	11,2
Higher general education (HAVO)	9,5
Intermediate general education (MAVO)	33,2
Elementary technical education (LTO)	11,4
Elementary nautical education (LNO)	0,2
Elementary domestic science education (LHNO)	9,2
Elementary agricultural education (LLO)	2,2
Elementary trademan's education (LMO)	1,2
Elementary commercial education (LEAO)	2,7
Combination HAVO-VWO	4,4
Other combinations	14,0

Table 1 : Schooltypes and enrolment percentage at grade level 8.(May 1981.)

In table 1 the major school types are given accompanied by the percentage of grade level 8 students who are in these school types. VWO, HAVO and MAVO are different streams in general education, while LTO, LNO, LHNO, LLO, LMO and LEAO are different streams within the elementary vocational education (LBO). In general students have different courses in each school type from grade level 7 in the Netherlands. But exceptions are possible, which are expressed by the "combination-types" displayed in table 1. Within these schools choosing for a specific school course is postponed until at least after grade level 8. The combination HAVO-VWO is the most common combination.

One of the major goals of the Second Mathematics Study in the Netherlands was to compare the implemented and attained curriculum between major school types. Because the mathematics courses in HAVO and VWO hardly differ at grade 8 level and beacause of the enrolment figures (see table 1) the population which was actually considered in the second mathematics study consisted of all students in the second year of HAVO/VWO, MAVO, LTO and LHNO. Using a stratified random sample of classes from this population, the study was conducted in May 1981.

Table 2 contains the numbers of teachers and students contributing to SIMS. The willingness of schools, teachers and students to cooperate was very high: about 98% of the distributed instruments were completed and returned.

	havo/vwo	mavo	lto	lhno	total
Teachers	60	70	57	49	236
Students	1515	1718	1276	991	5500

Table 2 : Number of participating teachers (equal to the number of schools) and the number of students.

For statistical reasons it was decided to sample a larger number of schools (and so teachers and students) in the two types of elementary vocational education, LTO and LHNO, than was needed to sample proportional to size. The larger numbers allow us to make precise estimations of variables for each of the school types in the project.

INSTRUMENTS.

The following tests and questionnaires were used:

1. Cognitive tests
2. student background questionnaires
3. teacher questionnaire 'opportunity to learn'
4. teacher background questionnaires
5. school questionnaire.

For this paper especially the instruments 1 and 3 are of importance.

The cognitive tests consist of 176 five-choice items. Each student answered 74 of the 176 items, by taking a test of 40 items, which was the same for all students (core test), and one of the four 34 item tests, each of which was designed for a quarter of the students (rotated forms).

In the 'Opportunity to Learn' questionnaire several questions are posed to investigate whether the subject matter, represented by the respective items, was taught to the students or not. In other words: did the students have an opportunity to learn the subject matter represented by that item? In the Netherlands, for each item teachers had to indicate in which of the following periods the subject matter concerned was or should be taught:

1. Primary school
2. 1st grade secondary school
3. 2nd grade secondary school: before Christmas
4. 2nd grade secondary school: after Christmas (but before date of data collection)
5. 2nd grade secondary education: after date of data collection
6. 3rd or higher grade secondary education
7. never.

To eliminate from this rating a hidden estimation of the difficulty of the item for a particular class, the teachter was also asked to estimate (per item) the percentage of students in his/her class who should be able to answer the item correctly without guessing.

RESULTS.

The Second Mathematics Study data bank contains data on various aspects of the mathematics curriculum especially for the second year of secondary education in the Netherlands. These data can be used in several ways: they give base line information for the year 1981, as well as the possibility of several exploratory data analyses which could result in generating hypotheses for future research. In this chapter we restrict ourselves to presenting some data on the actually implemented and attained mathematics curriculum in Dutch classrooms.

Number of weekly lessons in mathematics.

The Ministry of Education in the Netherlands does not prescribe the number of lessons per week (of 50 minutes) in mathematics for the second year of secondary education. In fact there are only regulations on the total number of mathematics lessons during the total duration of a school type e.g. in MAVO, which has a 4-year course, it is prescribed that the total number of weekly lessons of mathematics is at least 7. Schools are free in the way they spread these 7 or more lessons over the grade levels. For this reason it is interesting to describe the actual situation in grade level 8.
In figure 2 we see that the number of weekly mathematics lessons varies between as well as within school types. Although the mode in all schooltypes is 3 lessons per week, we see that the number of lessons slightly descreases from HAVO/VWO, MAVO, LTO to LHNO. The great variation between schools in LTO is striking.

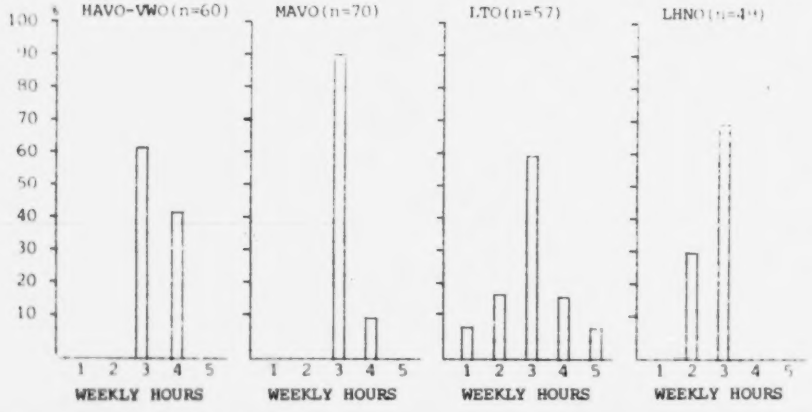


Figure 2 : Population A-distribution of the number of weekly hours in each school type.

Time devoted to mathematics topics.

As a consequence of the national examinations and their associated programs at the end of each school type and because of the use of the available mathematics textbooks, within a school type there is some uniformity in mathematics curricula in the Netherlands. On the other hand schools and teachers have much freedom to determine which mathematics topics will be taught at which time and with what emphasis in the mathematics classroom. Until now no systematically gathered information about the actual time devoted to mathematics topics has been available. To get an impression of the emphasis which is given to mathematics topics in the Second Mathematics Study teachers were asked to estimate the total number of hours during a year devoted to 14 mathematics topics. This kind of time estimation has some disadvantages. Firstly there is a possibility of overlap between the topics and secondly it is known that retrospective judgement of time allocation is not very reliable. But when we use these data only in relative rather than in absolute terms they are appropriate for description purposes.

Figure 3 shows some striking differences between the school types in the relative allocation of time to various mathematics topics.

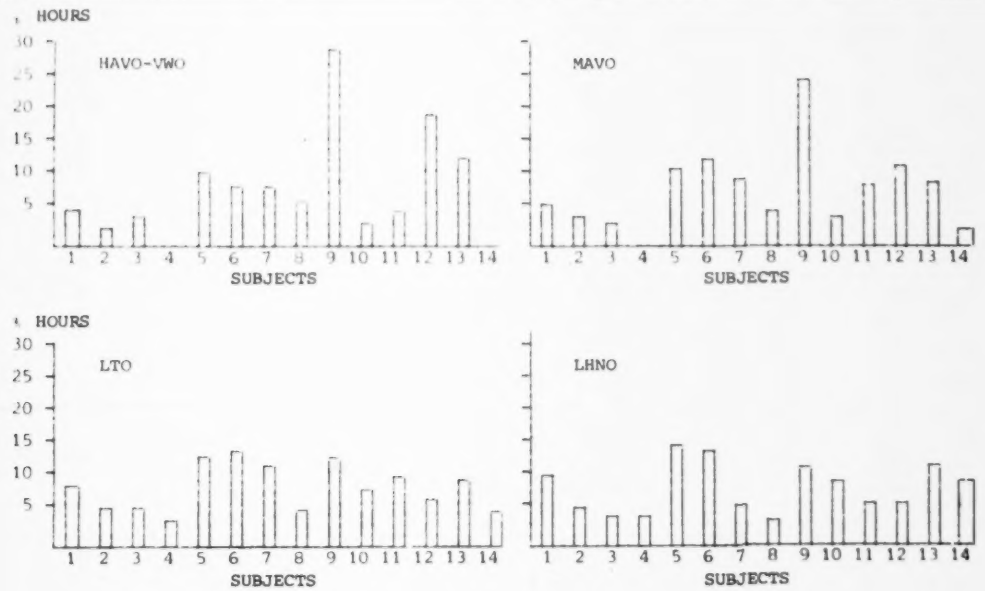


Figure 3 : Population A- relative allocation of hours for each mathematics subject in each school type.

In the general schooltypes (HAVO/VWO and MAVO) there is much emphasis on formulas and equations. In elementary vocational education (LTO and LHNO) the time is spread over many topics in comparison with HAVO/VWO and MAVO. Furthermore it becomes clear that the topics, probability and statistics, and percentage calculations have more emphasis in elementary vocational education than in general education.

The implemented curriculum.

In analyzing the opportunity to learn-data at the item level it becomes clear that within and also between the school types there is a large variation in teacher judgement of when mathematics subject matter is or was thaught. The results of aggregating these data to the forty core test items are given in figure 4.

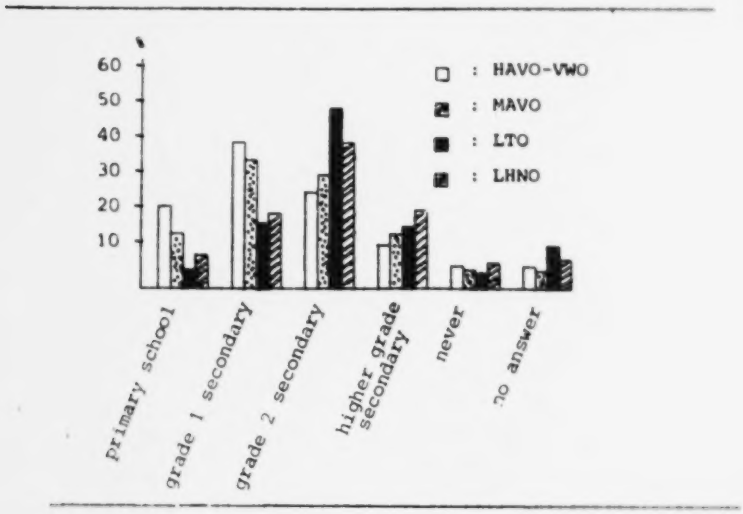


Figure 4 : Percentages answers (averaged over 40 core-items) from teachers to the question asking when the subject matter related to these items was taught to students in their class.

Notable in figure 4 is that the cognitive items fit the Dutch mathematics curriculum fairly well, because the percentages in the categories "never" and "no response" are relative low. Furthermore it is clear that in general education (HAVO/VWO and MAVO) mathematics subject matter is taught earlier than in elementary vocational education. Finally it is apparent that teachers in general education believe that more mathematics subject matter is taught in primary schools than do teachers in vocational education. This could mean that in vocational education quite a few primary school mathematics topics are repeated.

The attained curriculum (knowledge of students).

In table 3 the results on the mathematics tests are summarized. Means of percentages correct are given for all items and for five subtests.

SUBJECT	NUMBER ITEMS	HAVO-VWO (N=1486)	MAVO (N=1682)	LTO (N=1248)	LHNO (N=967)
Arithmetic (core)	11	82. %	63. %	48. %	36. %
Arithmetic (total)	19 à 21	80.	61.	47.	35.
Algebra (core)	9	86.	67.	46.	32.
Algebra (total)	16 à 18	79.	58.	40.	30.
Geometry (core)	11	77.	58.	48.	37.
Geometry (total)	19 à 21	74.	55.	45.	33.
Statistics (core)	4	91.	83.	69.	66.
Statistics (total)	7 à 8	85.	73.	61.	54.
Measurement (core)	5	80.	63.	53.	38.
Measurement (total)	9 à 10	79.	61.	54.	39.
Total (core+rotated)	74	78.	60.	47.	36.

Table 3 : Percentage correct answers for subtests and total test in each schooltype.

It can be seen in the table that the total scores and the scores on all the subtests, decrease from HAVO/VWO, MAVO, LTO to LHNO. This clear trend is not surprising, because it is known that the general abilities of students decrease in the same order in these school types. Furthermore this could be explained by the differences in time devoted to mathematics in these school types, as is shown in figure 2.

In the next section we will discuss in some detail the question of whether the actually implemented curriculum, as indicated by the opportunity-to-learn data, is related to the variation in the mathematics test scores. Before doing this some remarks will be made on the meaning of the opportunity to learn ratings.

VALIDITY OF OPPORTUNITY TO LEARN RATINGS.

In the second mathematics study mathematics teachers made the following judgements concerning all 176 test items:

1. Estimation of the percentage correct answers in the target class without guessing.
2. When is, or was, the mathematics necessary to answer the item correctly taught?

These judgements were made with respect to the mathematics class, that was involved in the study. The results presented in the following are based on the forty items in the core-test. A first analysis of the opportunity to learn results makes it clear that the two teacher judgements are not mixed up: the percentage items in the core test, which according to the teachter were taught before the testing date, has a low correlation (r=.25) with the mean of the estimations of percentage correct over all core test items. But from this result we do not know yet what meaning can be ascribed to the judgements. As a first exploration in this field we put the following questions:

1. What is the relation between the estimated and the actual percentages correct answers?
2. What is the validity of the judgements of whether the subject matter has been taught?

Concerning the first question it appears that the correlation between the actual and the estimated percentages is fairly high (r=.77) in the total sample. We suspect that the effect of differences between the school types is great, because the correlations within a school type are lower (in LHNO even 0). Nevertheless we can state that there is a strong relation in the heterogeneous total population. So we conclude that these are indications that the estimation of percentage correct is valid.

To answer the second validity question we use the following method:
we compare the judgements of teachers in the sample with data from other sources. From two other sources data are available on the period in which the mathematics needed to answer the test-item correctly was taught. First we have judgements from 4 experts from the Dutch testing institute (CITO).
These experts judged independently which test-items dealt with subject matter taught in primary schools. The other source of information is an analysis of mathematics textbooks conducted by experienced mathematics teachers. These teachers judged when the subject matter asked for in the test items was treated in the most commonly used textbooks in every school type. The primary school items in the core test were identified as follows: those items that the four experts unanimously judged primary school items. In figure 5 some of these items are printed as an illustration, while in table 4 the opportunity to learn results for all primary school items are given.

On the above scale the reading indicated by the arrow is between

A 51 and 52

B 57 and 58

C 60 and 62

D 64 and 66

E 66 and 68

Alexandra walked from Riverview to Bridgeport, which are 3.1 kilometers apart. During her walk she lost her watch, went back 1.7 kilometers to find it, and then continued in the original direction until she reached Bridgeport. How many kilometers had Alexandra walked altogether when she arrived at Bridgeport?

A 1.4

B 4.8

C 6.5

D 8.2

E None of these

The total area of the two triangles is

A 6×8 cm^2

B $\frac{6 \times 8}{2}$ cm^2

C $\frac{10 \times 6}{2}$ cm^2

D $\frac{16 \times 10}{2}$ cm^2

E $\frac{(6+8) \times 10}{2}$ cm^2

Figure 5: Some examples of items in the core test.

		Item														
		K03	K07	K11	K14	K15	K17	K18	K20	K21	K24	K26	K30	K33	K35	K37
havo/vwo	PRIM	63	23	42	15	22	18	62	27	37	3	45	58	77	53	25
	GR.1 SEC.	15	10	13	48	10	50	20	60	8	8	10	2	15	27	27
	GR.2 SEC.	10	7	38	30	5	28	12	8	5	72	25	3	2	10	12
	HIGHER SEC.	2	40	3	0	50	0	0	0	33	10	17	5	2	2	5
	NEVER	5	17	0	0	12	2	2	2	12	0	0	27	2	0	25
	N.A.	5	3	3	7	2	2	5	3	5	7	3	5	3	8	7
mavo	PRIM	56	11	26	19	7	24	53	20	17	1	44	60	51	23	20
	GR.1 SEC.	26	10	33	41	7	43	26	41	9	16	9	7	24	46	46
	GR.2 SEC.	9	17	34	34	23	29	16	33	17	71	14	7	13	26	14
	HIGHER SEC.	4	56	3	0	54	0	1	0	46	7	27	3	4	0	3
	NEVER	3	3	0	0	1	0	0	3	4	1	1	17	1	3	13
	G.A.	3	3	4	6	7	4	4	3	7	3	4	6	6	3	4
lto	PRIM	32	2	9	5	2	11	16	12	5	2	9	39	9	19	4
	GR.1 SEC.	16	14	26	26	9	25	37	33	7	18	14	21	26	28	28
	GR.2 SEC.	35	39	49	61	39	58	39	47	44	65	42	16	40	44	51
	HIGHER SEC.	5	37	4	0	37	0	0	0	37	9	28	0	12	0	7
	NEVER	0	0	0	0	5	0	0	0	2	0	0	12	0	0	4
	N.A.	12	9	12	7	9	7	9	7	5	7	7	12	12	9	5
lhno	PRIM	41	0	12	6	4	14	20	24	4	2	6	35	18	16	16
	GR.1 SEC.	27	18	22	12	12	14	35	43	8	0	14	18	16	51	27
	GR.2 SEC.	24	53	31	78	53	65	31	12	43	20	39	16	31	24	37
	HIGHER SEC.	2	22	22	0	27	0	2	8	39	57	33	0	27	2	6
	NEVER	0	2	0	0	0	0	0	0	0	14	0	20	0	2	10
	N.A.	6	4	12	4	4	6	12	12	6	6	8	10	8	4	4

PRIM = Primary education; GR.1 SEC. = Grade 1 secundary education; GR.2 SEC. = Grade 2 secundary education; HIGHER SEC. = Higher grade levels secundary education; N.A. = No Answer.

Table 4 : Percentage OTL-answers for primary education-items from the core test.

The rows in table 4 show for each school type the OTL-answer-categories. The table shows that in considering the total sample the OTL-instrument cannot be used for the identification of primary school mathematics. For the percentage answers in the category 'PRIM' are often low, whilst at the same time the percentage answers in the categories on secondary education are high. This might be realistic because many primary school topics in mathematics are repeated in secondary education. Looking at the shift of answers from the category PRIM in HAVO-VWO to the category secondary education in LHNO, this seems to be a plausible explanation, assuming that more repetition is necessary as fewer primary school goals are reached.

As far as the validity of the OTL-judgments is concerned, the following question might be asked: do teachers really teach what they say they do? In this study we only can answer this question indirectly. Direct answers could be given by performing observational studies, something that was not possible during SIMS. However, an indirect answer can be given by means of the textbook the teacher uses. A committee of teachers was asked to rate when the subject matter in the mathematics textbook, is taught. Each teacher was very familiar with the textbook they were asked to consider. Ratings were made for the following textbooks:

			Number of Raters
HAVO-VWO	:	moderne wiskunde	2
MAVO	:	getal & ruimte	2
		sigma	1
LTO	:	denken, doen en begrijpen	1
LHNO	:	passen & meten	1

These ratings can be compared with the ratings of teachers in the sample who use the same textbook. The textbook 'Passen & Meten' will be left out of consideration, because only 3 teachers in the sample used this textbook.

Table 5 shows how the committee of teachers rated for each textbook the occurrence of the subject matter necessary for answering a core-item before testing date.

1 = present in the textbook
0 = not present in the textbook
- = could be either

VWO: total 5 Wiskunde, 1 Sigma, MAVO: Moderne Wiskunde, DOB: Denken, Doen & Begrijpen.
CTBS: Percentage teachers in the study with the same textbook judging that the subject matter was taught in necessary circumstances.

Item	cAt		cAR		%		MW		DOB	
		(TBS)				(TBS)		(TBS)		(TBS)
1	1	100	1	1	100		1	100	1	100
2	1	37,5	0	0	9		1	100	0	23,8
3	0	12,5	-	-	18		-	47,3	0	52,4
4	1	100	1	1	100		1	100	0	52,4
5	1	100	1	1	100		1	94,8	1	76,2
6	1	89	1	1	64		1	78,9	0	28,5
7	0	37,5	0	0	9,1		0	21,1	0	61,9
8	1	62,5	1	0	36,4		1	97	1	81,8
9	1	100	1	1	100		1	100	0	28,6
10	1	100	1	1	100		1	100	0	52,4
11	1	87,5	1	0	9,1		1	78,4	0	33,3
12	1	87,5	1	1	90,9		1	94,7	1	57,1
13	1	100	1	0	18,2		1	93,8	1	81
14	1	100	1	1	100		1	84,2	1	95,3
15	0	47,5	0	0	9,1		0	15,8	0	28,6
16	1	100	1	1	100		1	100	1	89,5
17	1	100	1	1	81,8		0	73,7	0	94,2
18	1	37,5	-	1	63,6		0	47,4	0	52,4
19	1	25	1	0	0		0	52,6	0	14,3
20	1	100	1	1	81,8		0	63,1	0	66,6
21	0	50	0	0	9,1		0	5,3	0	19,1
22	1	100	1	1	100		1	100	1	81
23	1	100	1	1	100		1	84,3	1	47,6
24	1	87,5	-	1	81,8		1	89,4	1	95,2
25	1	87,5	0	0	46,4		1	73,7	1	33,3
26	1	75	1	0	18,2		0	42,3	0	23,8
27	0	25	0	0	36,4		0	21,1	0	28,6
28	1	100	1	1	90,9		1	100	1	90,5
29	1	100	1	1	100		1	89,4	1	61,9
30	0	0	0	0	9,1		0	0	0	14,4
31	0	37,5	0	0	9,1		0	15,8	1	19
32	0	62,5	1	0	0		1	36,9	1	13,8
33	0	50	-	0	36,4		0	15,8	0	23,8
34	1	75	1	0	81,8		1	100	1	23,8
35	0	75	-	1	81,8		1	57,9	1	66,7
36	1	87,5	1	1	81,8		1	87,5	1	33,3
37	1	87,5	1	0	74,7		1	52,7	1	71,5
38	1	100	1	1	100		1	100	1	85,7
39	1	75	0	0	0		1	47,4	1	14,3
40	1	100	1	1	81,8		1	79	1	100

Table 5 : Judgements for each core-item of the presence of the related subject matters in the textbook by members of the Teacher Committee).

Table 5 shows that the relation is fairly strong. Generally speaking, one can conclude that items for which there is no subject matter in the textbook are also less frequently taught. The contrary is also true. This shows that information from two different sources converges. The correlation between these two sources is .79.

Our conclusion from the preceding is that there are indications that the OTL-instrument is valid for the identification of the implemented curriculum in the first two years of secondary education. This means that the goal of measuring "Opportunity to learn" is to a reasonable degree, realised. Again it should be stressed that the instrument in this form is not suitable for the identification of primary school mathematics. According to our impression primary school mathematics can be identified in as far as it is not repeated in secondary education.

RELATION OF OPPORTUNITY TO LEARN WITH TEST-SCORES.

It seems reasonable to assume that along with other factors the presentation or non-presentation of subject matter will exert a strong influence on the knowledge of students. Students who are confronted with relevant subject matter in the classroom should - ceteris paribus - perform better than students who were not given the opportunity to learn the subject matter. In this section we will present a first analysis related to this topic.

First of all we investigated the relation between the amount of subject matter taught and the test-scores of students. Figure 6 shows the scattergram of these data. The conclusion is clear: there is no substantial relation, although the observed correlation is, due to a high N, statistically significant.

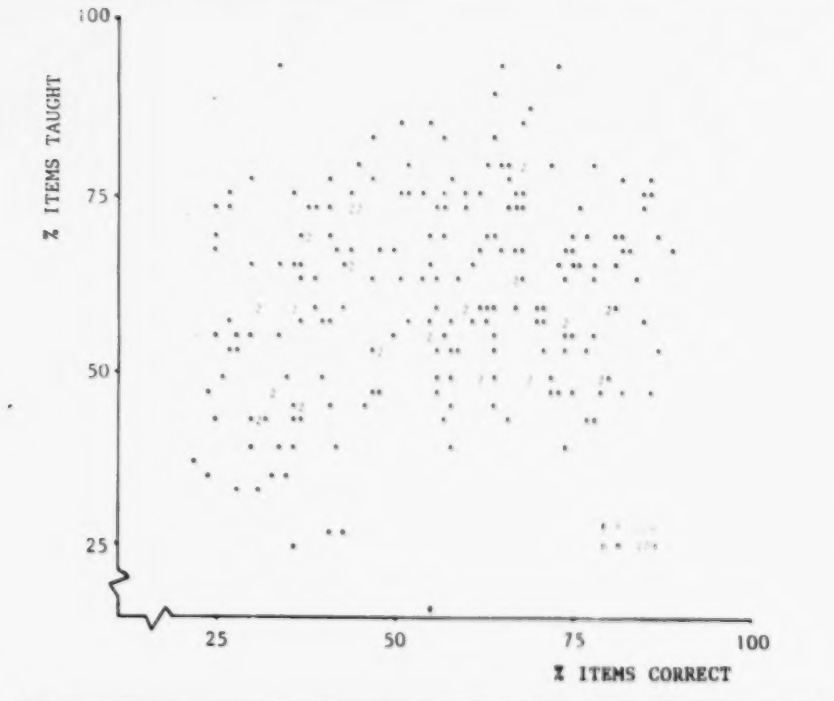


Figure 6 : Scattergram of the percentage items taught versus the percentage items correctly anwered in each class.

Figure 6 shows that the measure of Opportunity to Learn used in this study seems to be a bad predictor of student performance. This is a strange result because it would mean that presenting the subject matter has no effect. An alternative explanation might be that this result is due to the high level of aggregation of data (i.e. total test-scores). Analyses on the item level show that for many items there is an effect. This means that in such cases classes in which the subject matter is taught achieve much better than classes in which the subject matter is not taught. Results of these analyses will not be presented here. Ongoing analysis are necessary to try to explain why for some items the difference is large and for other items there is no difference.

CONCLUDING REMARKS.

In this paper some results of the Dutch participation in the Second Mathematics Study are presented. The validity of the Opportunity to Learn instrument was investigated and the data examined.

It was concluded that there are indications that the OTL-instrument is valid for the identification of subject matter which is taught in the first two years of secondary education. The weak relation of OTL with total test scores raises questions which should be answered in secondary analyses. Especially in this respect is the cross-national character of the study valuable, because comparable data from other countries are available. Ongoing analyses should investigate how possible different correlational patterns between countries can be explained.

As a result of these analyses we will hopefully in a few years know more about the quality of the OTL-instrument, which is especially interesting in curriculum implementation research.

REFERENCES

Eggen, Th., *Voortgangsrapport TWP* nr. 1, 1978 - juli 1979. Enschede: T.H. Twente, 1979.

Eggen, Th., *Voortgangsrapport TWP* nr. 2, juli 1979 - febr. 1980, Enschede: T.H. Twente, 1980.

Eggen, Th., *Voortgangsrapport TWP* nr. 3, febr. 1980 - aug. 1980, Enschede: T.H. Twente, 1980.

Eggen, Th., & Pelgrum, W.J., *Voortgangsrapport TWP* nr. 4, sept. 1980 - febr. 1981. Enschede: T.H. Twente, 1981.

Husén, T., *International study of achievement in Mathematics*. Stockholm: Almquist & Wiksell/Gebers Forlag AB., 1967.

International studies in evaluation, vol. 1-9, Stockholm; New York: Almquist & Wiksell; Wiley, 1973-1976.

Pelgrum, W.J., *Voortgangsrapport TWP* nr. 5, maart 1981 - aug. 1981. Enschede: T.H. Twente, 1981.

Pelgrum, W.J., *Voortgangsrapport TWP* nr. 6, sept. 1981 - febr. 1982. Enschede: T.H. Twente, 1982.

Sandbergen, S., *Nederlandse resultaten in tekstbergrip en natuurwetenschappen*. Amsterdam: Stichting Research Instituut voor de Toegepaste Psychologie, 1974.

Steiner, H.G., e.a., *Comparative studies of mathematics curricula - change and stability 1960-1980*. Bielefeld: Institut der Mathematik, 1980.

Treffers, A., *Wiskobas doelgericht*. Utrecht: Instituut voor Ontwikkeling van het Wiskunde Onderwijs, 1978.

Wiegersma, S., & Groen, M., *Resultaten van wiskunde-onderwijs*. Groningen: Wolters-Noordhoff, 1968.

Comment on the Dutch paper.

R.W. Phillipps
Department of Education
Wellington
New-Zealand.

I am grateful for the opportunity to comment on The Netherlands paper in this forum. I am especially pleased that the authors have seen fit to address the topic of the opportunity to learn variable and its validity in The Netherlands context.

Opportunity-to-learn has become one of the favourite sons of IEA since it was conceived in the first IEA Mathematics Study. In that study the variable asked the class teacher the approximate percentage of the students, 'to whom you teach mathematics and who are taking this set of tests', whether the topic in the question has been covered. The results from this question accounted for a substantiobal proportion of the variance in a number of the countries. I note though that The Netherlands apparently did not administer this question in the first Study. In the Science Study the measure was refined although in this instance it was the comnination of all the opinions of the science teachers in the school that formed the variable. In the same study OTL entered the regression model in most countries.

As I have said earlier, in the second mathematics study the prime aim of the design was to address the problems of mathematics educators. Considerable thought was given to the form the OTL variable should take. A number of the obvious problems with the variable were carefully considered. It was clear that we would need a measure which applied to the class and hence the question should be addressed to the individual class teacher. It was also suspected that in many cases the teachers' response was conditioned by their perception of how many in the class would actually get a particular item correct. In the hope of eliminating this contusion of thinking, the question 'What percentage will get the item correct?' was asked directly before questions to investigate whether or not the mathematics behind the item had actually been taught (or had been assumed as taught) to the teacher's class. The Netherlands translation of this question has had the effect of their being able to locate accurately the time when the teaching event took place. I would hope in subsequent papers the authors might follow up this aspect of when the item was taught in respect to the testing date.

Before looking at the results in the paper might I just say a little in general about the interpreation of the results from this study. Despite the obvious interest many countries will have in their international ranking, I do hope that cognitive outcomes do not get displayed without a number of caveats. It is clear that the cognitive tests are not entirely appropriate for all countries.

Still it is hoped that in this study the international tests match the Intended Curriculum in all countries better than in the First Study. The degree of appropriateness for a country is essential information. Even though a mathematics topic might appear in an official syllabus, if the students' clearly have not had an opportunity to learn it, their chances of a correct answer must be very low. Even in their simplest forms the appropriateness ratings and the OTL measures can signal warnings to any reader tempted to take a set of cognitive scores at its face value. The interesting questions for each country are the ways in which the Intended, the Implemented and the Achieved curricula mesh. I hope this will be kept in mind in all National and International reports. The emphasis being placed on the OTL variable in The Netherlands to my mind bodes well for such responsible reporting. I would commend it to all countries.

One other caveat that must apply to all those countries which wish to make comparisons between the first and second studies, is the need to carefully consider the comparibility of the samples particularly where different retention rates apply at the different times.

Perhaps talking about samples takes me back to the paper. The authors have given a clear indication of the school types they have included in their definition of Population A. However, I feel in an international forum it is unlikely that foreign readers would appreciate that some 20% of the internationally defined population has been excluded from The Netherlands population - another caveat that should appear alongside any result of outcomes which are likely to be generalised to the country level.

As one of the 'alterable' variables, time spent on mathematics, is obviously important both within and between countries. Despite the need to rely on retrospective teacher judgements of the time spent on various mathematics topics the data displayed in the paper showing the relative allocation of hours to 14 mathematics topics should be of intense interst to the policy makers and curriculum planners. I am intrigued to know how well the pattern displayed for the 4 school types matches the perceptions of what was intended by the official curricula. If they do not match is this a subtle subversion by the textbook writers?

From the efforts made by The Netherlands to validate the OTL responses through the use of a group of teachers analysing the most popular textbooks I presume the textbook occupies an important role in mathematics teaching in The Netherlands - as I suspect in all countries. I hope all countries will follow The Netherlands in analysing their textbooks. This validation exercise I think has unearthed a rather universal factor - that no matter how well qualified or skilled teachers may be, their knowledge and capacity to judge exactly what mathematics has been taught in previous years is suspect.

For population A in The Netherlands this may not upset the OTL judgements quite as seriously as in New Zealand where Population A is the first year in the secondary system. At least in The Netherlands the teachers were aware of the previous year's curricula. In New Zealand I fear the disjunction of knowledge could be having quite traumatic effects for may students. In itself this issue must be of concern to all planners concerned with the unity of the school curriculum. I would be interested in learning of any attempt in The Netherlands to bridge the curriculum gap between the primary and secondary schools in mathematics.

I sense a little disappointment from the authors that having accepted at least a limited validity for the OTL measure they found that no real clear relationship exists between the OTL and the class cognitive achievement results in The Netherlands. Perhaps we are asking too much of the OTL measure when we look for strong correlations between OTL and individual student scores.

Clearly teachers can honestly claim they have gone through the motions of teaching the mathematics required to answer the item. Only analysis at the classroom level could throw light on how succesful the teacher had been in teaching all students in the class. Hence the greatest strength of the measure may well lie in the between country analyses. It could tell us whether or not teachers are attempting to implement the official curriculum. A similar investigation might welll be undertaken between school types in The Netherlands. If there are countries with strong correlations between OTL and achievement, it could well be telling us something about the effectiveness of a country's teachers' ability to reduce within class differences.

The attached paper by R.A. Garden, International Coordinator for the Study, sets out additional reasons why weak correlations are not unexpected. However, it is early days yet and there are still many ways the opportunity to learn variable could be explored. I am sure that the group an Enschede will have already considered analysis by subscores which I would predict would produce stronger relationships. Would it not be worthwhile to explore the responses by teacher qualifications? An indication from the New Zealand results is that the teacher's estimate of how many students in the class will get an item correct may turn out a better predictor of achievement - this measure must contain some element of the opportunity the teachers believe have been offered the class as well as their perception of their effectiveness. It could well be explored in this context.

Whatever the final outcome of the national and international investigations of this variable, I am sure its importance will not be underestimated and once again may I plead for at least sounding the need to consider the concept of OTL when anyone mentions a cognitive score.

ATTACHMENT: Teacher opportunity-to-learn and achievement (by R.A. Garden).

Several investigators have expressed surprise and dissapointment that the relationship between teacher judgement of opportunity to learn and achievement is not generally a strong one. The follwing outlines some of the reasons for the expectation of a strong relationship in the IEA Mathematics Study being unrealistic.

Minumum prerequisites for a strong relationship include:

1. Accuracy of teacher judgement of OTL.

   This depends on the teacher knowing well what is prior knowledge, absence patterns for the class and also calls for rather difficult judgements for some items. Consider the item $(-\frac{1}{4}) + (-\frac{1}{4}) = ?$ If a class had had the opportunity to learn items like $(-5) \cdot (-2) = ?$ and $\frac{1}{4} + \frac{1}{2} = ?$ the decision about whether the class had had the opportunity to learn the mathematics needed to answer $(-\frac{1}{4}) + (-\frac{1}{4}) = ?$ correctly is not straightforward. For high ability students the answer is probably "yes" and for low ability students "no". Much finer shades of judgement are needed for items which test skills which have been taught but which are not quite in a form students are familiar with. TOTL scores tend to be low for analysis level items, for example, even though the skills required might only be addition or substraction of whole numbers and the recognition of a pattern.
   Indications are that teacger judgements tend to be reasonably accurate, especially when aggregated, but reliability would be a long way from 1.

2. Item sensitivity to OTL.

   Items must be such that they will be answered correctly by almost all students judged to have had a opportunity to learn the required mathematics and incorrectly by those judged not to have had this opportunity. IEA items were not selected with this as a criterion. Items with high discrimination were selected but discrimination was based on general mathematical ability, not on OTL. Thus for an item on which the TOTL measure is 100% it can be expected that better students will give correct answers and weaker students incorrect answers. For maximum discrimination on IEA criteria between 50% and 60% of students will answer correctly.

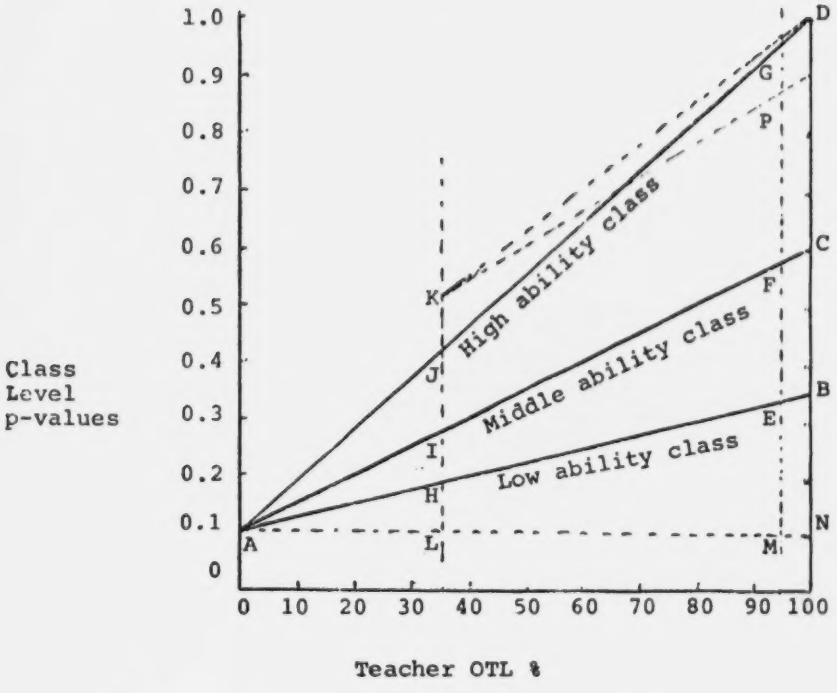
3. Negligible decay (fade) of learning.

   Some items of test matter will have been learned up to 3 or 4 years prior to the target year in some countries. For these items students are judged to have had the opportunity to learn. In many cases they have also had the opportunity to forget. Thus an item which has a 100% TOTL measure will produce a range of mean p-values across classrooms.

Studies of OTL by ETS found that for strong relationships between OTL and achievement items should test material that has been recently taught and should be at the computation level of behaviour.
Consider the attached figure. In it some relationships between teacher OTL and class level (mean) p-values are hypothesized.
Given the nature of the items, ie, 5 alternative multiple-choice very few will have mean class p-values below 0.1 (line AN). AD then represents the ideal, with the mean p-value approaching 1 as OTL approaches 100%. This would happen with classes of students who were of very high ability but yet were ubable to solve problems containing untaught matter by building on or making deductions drom knowledge they should have. Experience with the "French" items at population A and population B levels indicates that the very able students are indeed

able to do this. Thus KD represents an conservative upper limit for mean p-values for classes of able students and allowing for careless responses, misreading of questions etc KP might be even more realistic.
Item selection methods ensure that there also effective upper limits for "middle ability" and "lower ability" classes and that these are likely to be of the order shown by lines AC en AB respectively. Items which were too "easy" (ie likely to be answered correctly by more than 80% of students) were rejected to avoid ceiling effects unless there were other good reasons for retaining them. A "good" item for the purposes of the study would have a p-value for a national sample of around 0.6 and thus middle ability classes would have mean p-values at about this level. For similar reasons lower ability classes have an effective upper limit to mean p-values as shown.
Very few countries have more than a handful of item OTL values below 35%. KL then represents a boundary on the left of which very few points on a plot of OTL against mean class p-value would fall. For a number of countries few items have OTL measures above 95% so few points would fall to the right of MG. We are left with the area LMPK within which almost all points would fall. Since there is, in most countries, a continuum of class ability levels points will occur throughout the region. If all the points were within this region, and if, as is likely, points were rather symmetrically placed with respect to the line of regression, low correlations could be expected. This is especially true for subsets of items for which KL is much further to the right than in the figure.
So why do we get moderate correlations for some countries on some subtests? More by accident than design there are a few items which fall near A in the region ALJ. These items, it seems to me, help to "fix" the regression line and by increasing the OTL range while at the same time being confined by a decreased p-value range, to raise the correlation coefficient.
In the case of The Netherlands there are effective lower limits to both p-values and OTL measures for each school type. The effect of this is probably to spread point widely in the region LMPK and thus produce a low correlation. It is likely that if OTL and p-values were aggregated to school type level the points would be well ordered and the correlation high.
If the above discussion has any validity a stronger realtionship between TOTL and achievement should be obtained if class ability level is controlled. As a rough and ready check of this proposition Teacher OTL measures aggregated over all core test items for each NZ class were correlated with class means. With all 199 classes included the correlation was 0.23 while with the 30 highest and 30 lowest scoring classes excluded the correlation was 0.30.
For The Netherlands data item level TOTL (aggregated to country level) was correlated with p-values of all computation level items in the test forms for Population A. The correlation obtained was 0.42.
Strongest relationships should be expected for content subtests of computation level items, especially for those subtests where most of the content was taught in the target year. At this point, however, the aim of the exercise must be considered. Any subtest which fulfils all of the above conditions for high correlation between TOTL and achievement is unlikely to be adequate as a criterion variable for the broader purposes of the study. The utility of TOTL as a key predictor variable in a causal model is therefore yet to be determined.

HYPOTHETICAL P-VALUE BY TOTL



Section III

The International Study of Achievement in Written Composition.

General Overview*

A. Purves and S. Takala
University of Illinois
Urbana, Champaign
U.S.A.

The main criterion variables for this study are (1) students' performance on a number of writing tasks and (2) students' attitudes toward schooling and toward composition writing. The main explanatory variables deal with descriptive characteristics of the school and its community, the educational program and curriculum of the school, teaching practices, students' home background and interactions in the home, student motivation, and the amount of writing undertaken in class and in contexts outside the school.

The instruments were developed mainly by staff at the Coordinating Center for the study in collaboration with the Steering Committee whose members represent five different countries. Before the first drafts were prepared, theoretical analyses of problems related to them were carried out. Several alternative approaches of carrying out the study were discussed before a selection was made by the International Study Committee (ISC).

Cognitive variables.

The selection of the writing tasks, which are the key cognitive criterion variables, was based on a number of considerations. National Centers provided information about the curriculum on a detailed questionnaire. They also sent copies of important examinations in written composition, and lists of typical writing assignments. In addition to this information, which was collected to maximize curricular validity, the selection of tasks was also based on a theoretical model of the domain of writing and of written composition, developed by the Steering Committee. Like all preparatory work done by the Coordinating Center and by the Steering Committee, the tasks were checked for relevance and appropriateness by the National Centers, whose comments were used in the revision of instruments.

There has been a high degree of agreement on the theoretical model of the study, and on the model of domain specification. It has been more difficult to obtain an equally high agreement on the particular selection of tasks from the domain. On most tasks the agreement has been high. Where disagreement has occurred, it has concerned the emphasis of different types of tasks. A few National Centers felt that narrative/descriptive/expository tasks were unduly emphasized at the expense of more expressive writing. New tasks have been developed to take this into account. Subject to the decisions of the International Project Council (IPC) and the ISC, they may be included in the international core component or may be made available as national options.

Explanatory variables.

The first drafts of instruments to measure the explanatory variables were based on the model designed for the study. In revising the school, teacher, and student questionnaires prior to pilot-testing, the Steering Committee drew upon the model and on the comments from the National Centers.

The appropriateness and clarity of the instruments is being pilot-tested in all countries participating in the study. Pilot-testing data will be used in revising all the instruments for the main testing program.

It can be seen that the development of the instruments is based on the cooperation of all participating countries. The Coordinating Center and the Steering Committee usually provided the initial plans and drafts for the National Centers, which in turn provided feedback to assist the Steering Committee to revise the instruments.

* Reprinted from IEA-Newsletter, no. 2, july 1982.

Results and effects of IEA Written Composition Study in The Netherlands.

H. Wesdorp
University of Amsterdam
Centre for Educational Research
The Netherlands

INTRODUCTION

The IEA written composition study has three aims:
1. Description of the written composition curriculum in various participating countries.
2. Description of societal, school and individual variables that might influence achievement in written composition.
3. Analysis of the relations between achievement in written composition and the instructional variables in the light of the background variables (societal, school and individual).

In Holland - and probably also in the other participating countries - we started by making preparations for the description of the status quo. In the first place we carried out research into the literature in order to establish what instructional variables might be important in written composition teaching. We presupposed that an overview of the empirical research into the effectiveness of various instructional variables would provide us with a number of variables which we should have to include in our descriptions. We also assumed that it would in any case be no bad thing to gather together the empirical research results of several decades in a clear and readily comprehensible form. Particularly in a country like Holland, where little or no attention has been explicitly directed, by means of systematic research, at the charting of the process of writing instruction, a review of this kind would be anything but a luxury. I shall shortly present some of the results of this literature review in a strongly condensed form.

In the second place we drew up a questionnaire aimed at the most important aspects of the instruction process and sent it to a representative sample of Dutch secondary schools, where they were filled in by teachers of the first, second, third and fourth forms. We are now, for the first time, able to say how written composition is taught in Dutch schools. Our results have so far only been partially worked out; I shall shortly tell you about some findings which are based on some 700 completed teacher questionnaires. We hope soon to publish a report containing a detailed analysis of the survey results, based on more than 1000 teacher responses.

In the third place, besides the results of the first phase (i.e. the literature review and the description of the status quo in the classroom) I can report to you on some side-effects of the IEA written composition study. It is not just the research results themselves that are important: the effects that they have on the work of other investigators can also count as part of a project's 'yield'. And in our institute the literature review led to the idea of introdu-

cing and testing in the Dutch situation a number of instructional techniques that experience abroad had shown to be interesting. The first such project, scheduled to last three years, has already started. It is designed to study the effects of peer evaluation on the quality of pupils' written work.
A second project, also of three years, is now in preparation and will examine how certain prewriting activities, particularly training in structural and organizational skills, influence the quality of written products. I shall be saying more on this subject in a moment.

AN EXAMINATION OF THE LITERATURE TO ESTABLISH CRITICAL VARIABLES IN WRITTEN COMPOSITION INSTRUCTION

We examined 158 quasi-experiments on the effectiveness of various instructional variables. We distinguished 18 different instructional techniques. Some of these, such as teacher feedback (i.e. various ways in which the teacher can provide feedback) proved to have been studied with some frequency, others less often. Here I shall review only the most important results of this research overview (Wesdorp, 1983). It is important to know which instructional variables have clearly positive effects on pupils' written composition and which do not. The chief results are these:

Clearly positive effects are shown to be the result of various pre-writing activities, i.e. activities focusing student attention on the question of how to organize a text, how to generate ideas, on problem-solving categories, and on rules for systematically treating a subject. Among the pre-writing activities, the pre-writing discussion about the composition task is the most wellknown and this task receives so much attention that a special paragraph has been devoted to it. Positive results are also achieved by using stimuli (writing assignments) that fit the individual student. Of course, the question remains whether education must always aim for writing tasks that 'fit' the students. It is, of course, possible that the educational goals may not match the general preference of the students. Nevertheless, the fact of the positive results of investigations using stimuli chosen by or personally experienced by the student is a factor that should not be discounted in the teaching of composition. Transformational sentence-combining exercised also appear to have a positive influence on writing ability. In a fairly large number of experiments such exercises have not only turned out to show an abundantly clear positive effect on a student's sentence structure - a result that was to be expected - but also on the general quality of the written material.
Clearly positive results were also found in experiments in which the effects of peer-evaluation were studied. The great practical advantage of this approach, which lightens the extensive evaluation task of the teacher, is obvious. But apart from that, the effect of this approach on writing abilities are, to a large extent, positive. This effect may be explained by educational and communicative theories. The revision of a text, i.e. rewriting after comments by the teacher, also has shown positive effects on the quality of the written material.

Doubtful effects. Most of the experimental variables discussed in the literature do not show very convincing results. In the majority of experiments, a fairly large group of instructional variables do not show positive results. This does not prove conclusively that these instructional variables are irrelevant. It does mean, though, that the positive expectations that were held on theoretical grounds have not been fulfilled in the course of the investigation. Or to put it differently, there is not much evidence, so far, to favor these experimental variables. Of course, the validity of such a statement depends on the, sometimes questionable research designs.

What are the experimental variables with doubtful effects? They are: the

workshop/writing-lab approach, group work, the reading models approach, the methods that lean heavily on individual conferences between teacher and student or tutor and student, the individualized approaches, and the approach in which the writing frequency is stepped up without special assistance.

Limited or no effects. Finally, there are a number of instructional variables for which research has conclusively shown that they have very limited or no influence on composition ability, despite originally high expectations. For example, the general effect of various forms of teacher-evaluation has been disappointing. The way in which the teacher gives the evaluation (positive or negative and corrective, extensive or concise, detailed or global) seems to make little difference. The group of investigations on teacher-evaluation does not offer us a clear direction. The question, 'which form of teacher-feedback works best?' remains unanswered until now.

Equally lacking in convincing results are investigations into the effects of various approaches based on grammar. It turns out that the 'traditional grammar approach' clearly has little effect in comparison with approaches advocating a more direct training of language abilities. The use of 'structural grammar' does not show convincing results either, as is the case with the use of 'transformational generative grammar'.

THE PRESENT STATE OF WRITTEN COMPOSITION INSTRUCTION IN HOLLAND

We asked teachers of Dutch schools questions about their objectives in writing and about the emphasis they place on particular aspects and exercises. We also asked them about the way they teach written composition in the narrower sense: the kind of asssignments they give their pupils, the teaching material used and the way pupils receive feedback, and the amount of time spent on the subject.

Here are some of the provisional results. These are valid for secondary education as a whole, though naturally there are differences between the various kinds of secondary schooling. Our sample embraced the chief types of school, i.e. both the (lower) vocational types of school and the (higher) more academic types. Moreover there are also, of course, differences between classes; we have put together the results of the bottom four classes in all the schools, so that the sum total is an overall, approximate, and - because these results are taken from only about two-thirds of all the questionnaires returned - provisional result.

Objectives.

A clear majority has the opinion that the general purpose of our teaching of written composition is to enable the pupil to communicate well in a variety of practical situations (e.g. at home or at work). This objective, oriented on practical communication, is more highly favoured than other objectives which have more to do with the pupil's personal or intellectual development.

Particular aspects and exercises.

This practical objective aimed at by the majority of teachers is in sharp contrast to what they actually do in the classroom. Much attention is paid to semantic exercises (vocabulary and the correct use of words), spelling and writing conventions, parsing and naming the parts of speech; much less time is devoted to 'practical' exercises in collecting and selecting from material, constructing an argument or the systematic treatment of a subject. If we look at the sorts of assignments that are popular in written composition teaching in Holland, we are forced to the same conclusion: the emphasis is certainly not on practical communication.

The various sorts of letter-writing activity are unpopular, and the same goes

for instructions, announcements, circulars, short notes or advertisements. The most popular sorts of assignment are narrative topics: reproducing a story heard, completing a story already started, writing a personal narrative, a report in narrative form, or a personal anecdote. This attention to personal and narrative forms of writing is accompanied by attention to summaries and book reviews.

Teaching material and the design of written composition teaching.

The majority of teachers use textbooks as their teaching material, though 'home-made' material or material collected by the teacher is also used.
The way written composition is taught is determined largely by the individual teacher. Generally it is the teacher who decides what subject the pupils are to write about. Sometimes the pupils have a choise from a number of set subjects, but very often they are then not allowed to decide what sort of writing task they will perform using the stimulus material provided (which is generally a short list of 'titles').
Composition is individual work, with every pupil working on his own: not much group work is involved. Although teachers say they pay attention to the various stages of the writing process, very few of them get the pupils to collect information, have brainstorming sessions, draw up their own guidelines, or revise their own texts. A minority link written composition as practically as possible to realistic situations with a real purpose and an actual audience.

Feedback.

How do Dutch teachers provide their pupils with feedback? The vast majority make comments and suggestions for improvement in written form: each pupil is given back his essay, with assessment and/or corrections, to read for himself. In quite a lot of cases the most important mistakes are dealt with and discussed in class. What form do teacher comments usually take? Most commonly one finds the following variants: the teacher gives a numerical or alphabetical mark, or a single word with no further comment; he provides a written comment of two or three lines at the end of the essay; or he makes more detailed comments, including suggestions for improvement. But what happens with these comments is uncertain: there is no evidence to show that teachers make a habit of having their pupils revise their written work on the basis of the comments passed on it.

Besides teachers, pupils themselves are also involved in the feedback process - albeit to a much lesser extent. Fifty per cent sometimes, and fifteen per cent often get their peers to comment on their work.

Time spent.

How do teachers divide up their time on the various aspects of verbal ability? There are considerable differences between individuals, but on average something over 2 teaching periods is spent on writing per month, compared with 3 on reading. Thus written language skills receive something over 5 periods a month. The same applies to oral skills: speaking (more than 2), listening (just under 2) and discussing (1) also add up to about 5 periods. Besides the two periods in class (approximately 1½ actual hours), the same amount of time is devoted to writing as homework. The amount of time spent on writing is probably low: if one compares the two teaching periods a month with the total number of teaching periods (about 120) one has a telling illustration of how important we in Holland regard the teaching of written skills of self-expression.

I have now covered some provisional results from our examination of the status quo - research which has for the first time provided something like an accurate picture of written composition teaching in Holland. We are also preparing

an overview of the theoretical discussion that has taken place in Holland over the last few decades concerning the didactics of teaching written composition. For foreigners, of course, this theoretical discussion is probably a good deal more interesting than the somewhat hazy (not to mention grey!) picture of reality that I have shown you. Unfortunately at present our survey of the chief schools of thought about the teaching of written composition is not yet finished. It will be published in the very near future with a summary in English (Damhuis, De Glopper & Wesdorp, 1983).

SIDE EFFECTS OF THE IEA WRITTEN COMPOSITION STUDY

One of the arguments leading to the start of the study was that it would highlight the importance of written composition skill itself and would also stimulate other research. It has certainly done that: the review of empirical research, viz. quasi-experiments which first and foremost set out to study didactic variables manipulated by the teacher. The criticism that this sort of research has attracted is well known: it starts from a 'scientific' model, ignores many contextual factors, and fails to do justice to what is going on inside the pupil himself. We therefore also reviewed the research that has concentrated on the writing process since the early seventies (Bochardt, 1983).
There are also other signs that interest in the increasing body of literature on the writing process is growing in Holland. The research review also draws attention to the possibilities of peer evaluation as an educational principle, and to the positive aspects of certain prewriting activities in the field of organizational skills.

A project to investigate the possibilities and effects of peer evaluations has recently started (Rijlaarsdam, 1983). It is designed to establish whether having written material read and assessed on a regular and systematic basis by fellow pupils has any effect on a pupil's own written composition skills. The project allows such effects to be detected in two ways. In the first place by examining differences in the writing process, particularly in planning and revision behaviour, and second by detecting differences in aspects of the end product, in particular its audience orientation. The project started in April 1983 and is scheduled to run until 1986.

Another project, designed to study the effects of training in structuring and organizational skills on aspects of written composition ability is currently in preparation (De Glopper, 1983). Here the research carries on from the quite positive results of literature studies in the field of prewriting activities. One group of pupils will learn to use heuristics (problem-solving procedures) for solving structuring and organizational problems arising during the writing of texts of an expository or explanatory nature (expository writing). Another group will receive theoretical instruction on textual structure and organization, and will be taught to analyse the structure and organization of model texts. The project, which is due to take about 3½ years, will look in various ways at the effects of these two forms of prewriting activity (De Glopper, 1983). It is scheduled to start in the spring of 1984.

CONCLUSIONS

In Holland the IEA Written Composition Study has produced results that are both valuable and new: an overview of the literature and a review of the status quo which for the first time equips us with a good idea of how, and to what extent, written composition is currently taught at Dutch secondary schools. At the same time the IEA study has provided the impetus for an in-depth study of the

literature on the writing process, and for exploration of the possibilities of peer evaluation and certain prewriting activities. Moreover, for many of the participants the international co-operation between mother-tongue specialists are not a little nationally oriented, which has meant that the exchange of knowledge acquired in this field has been slower than it might have been. The fact that the IEA Written Composition Study may lead to the publication of an 'International Review on Mother Tongue Education', the first truly international journal in the field, testifies to how stimulating working in an international environment can be.

REFERENCES

Bochardt, I.: *Onderzoek naar het schrijf- en planningsproces in de zeventiger en begin tachtiger jaren; op weg naar een theorie.* Amsterdam, SCO, 1983.

De Glopper, C.M.: *Onderzoek naar effecten van training in opbouw- en organisatievaardigheden op aspecten van de stelvaardigheid. Onderzoeksplan.* Amsterdam, SCO, 1983.

Damhuis, R.C.M., De Glopper, C.M. & Wesdorp, H.: *Schrijven in het voortgezet onderwijs. De theorie en de praktijk van het stelonderwijs.* Amsterdam, SCO, 1983.

Rijlaarsdam, G.: Audience-oriented writing and peer evaluation. A paper presented at the CCCC Conference, Detroit, March, 1983. Amsterdam, *SCO-cahier*, 1983.

Wesdorp, H.: *Schrijven in het voortgezet onderwijs. Een overzicht van het onderzoek naar de effecten van diverse instructievariabelen op de stelvaardigheid.* Stichting voor Onderzoek van het Onderwijs, Harlingen, Flevodruk, 1983.

Comment on the Dutch paper.

A.C. Purves
University of Illinois
Urbana - Champaign
U.S.A.

Dr. Wesdorp has described certain results of the IEA Study of Written Composition for The Netherlands. As Chairman of the International Project Council for Written Composition, I am pleased that these early results are precisely what we hoped would be the outcomes of the initial phases of the study for each participating country, and I can report some similar results from elsewhere.

The IEA Study of Written Composition, like many other of the IEA studies, has two broad aims. The first and announced aim is to provide descriptive data concerning the performance of students in a school subject and to relate their performance to data concerning school policies and practices. By developing a study cooperatively across nations, the data derived provide policy makers and teacher educators with the possibility of exploring alternatives to the current practices of a country. They also can provide a lens through which a policy maker or teacher trainer can see a particular country's results as a choice rather than as a necessity. The finding, for example, that students in one country write excellent narrative compositions but poor argumentative compositions and that the students in another country write excellent arguments and poor narratives tells the educator that good argumentative writing may result not from some developmental law but from a curricular decision in each of the countries.

The second aim of an IEA study is clearly as important as the first. The Writing Study, like previous IEA projects, enables a group of subject matter specialists from around the world to work together to define the domain and the various alternative practices in instruction in that domain. In many cases, the domain had not been fully conceptualized before IEA came onto the scene. Such was clearly the case with literature and civic education. So too, it has been with written composition.

During the first three years of the project, the members of the Project Steering Committee have had to undertake a number of tasks. The first was to define the domain of school writing internationally. The second was to define the major constructs in the pedagogy of writing, that is, those practices and strategies that seemed to differentiate writing curricula and to affect student performance. The third was to define achievement in written composition from an international perspective - so as to give instruction to those who would score the students' compositions. The Committee has accomplished these tasks and in so doing has involved the National Research Coordinators from each country.

A part of the results of this work appears in the volume published by Pergamon Press in December 1982, "An international Perspective on the Evaluation of Written Composition". Portions of this volume have also been published and made available to teachers in Finland, Italy, The Netherlands, Hungary, and the United States. A second publication is that of Dr. Wesdorp, which has, through

its abstract, received broad attention in the United States. In Italy, the Teacher Questionnaire, because of its detailed exploration of alternative instructional practices, has been used as the basis for in-service education programs. In Indonesia, the study of the background for the curriculum has produced one doctoral dissertation. The scoring scheme that has been developed has now been tried in other projects and seems on its way to affecting large scale assessments in several countries, even some not directly participating.

In a sense, these effects of an IEA study are fugitive, but they are perhaps the most long-ranging effects. I believe that in nearly every field that it has explored, the IEA method of cooperative inquiry across nations and across languages has brought about subtle and profound changes in the way people think of a subject area, of curriculum and instruction in that area, and of the assessment of student performance in that area. Such effects as those Dr. Wesdorp has described for the Netherlands are as important as, if not more important than, the final volume of a study.

Section IV

The Classroom Environment Study.

General Overview*

B. Avalos
Ontario Institute for Studies in Education
Toronto
Canada

In most countries of the world the predominant form of teaching still involves a single teacher engaged in 'classroom' teaching and learning with a group of about 15 to 50 students.

Although other patterns of organization are possible, it would be difficult to develop any mahor changes in these well-established institutions and practices.

It follows that a faster method of improving the education of students in schools across the world is to acknowledge the problems associated with major organizational changes and, instead, to focus on improving the quality of the education taking place in traditional classroom situations.

Working on the above assumption, the IEA Classroom Environment Study: Teaching for Learning represents a collaborative effort to identify classroom processes and factors that affect student learning and to develop teacher-training programs that are grounded on empirical research.

AIMS OF STUDY.

The general aims of the study are:

- to identify teaching practices which are correlated with improved student achievement and attitudes;
- to examine the relationship between such teaching practices and both contextual factors and student learning behaviours;
- to determine the degree to which those teaching practices can be fostered through relatively simple teacher-training programs;
- to determine the degree to which the training and the changed practices cause improved student achievement and attitudes.

In pursuing its aims, the IEA Classroom Environment Study has been designed as a two-stage effort covering a period of five years. In the first stage - the Correlational Study - an attempt is being made to examine the relationships between contextual factors, teaching practices and student learning behaviours and to identify teaching practices correlated with improved student achievement and attitudes.

The major alterable variables being examined at this stage are: time on task, feedback, correctives, cues and questioning. The correlational findings are then to be translated into recommendations for teaching practices to be used in the development of teacher-training materials and programs.

In the second stage - the Experimental Study - the teacher-training programs based on the results from the first stage will be given to an experimental group of teachers. Studies will then be conducted to determine the degree to which the recommended teaching practices have been fostered through the training programs, and the degree to which the training and the changed practices contribute to improved student achievement and attitudes.

INSTRUMENTS AND DATA COLLECTION.

At present twelve countries at various stages of industrial development are participating or planning to participate in the Correlational Study during 1981 - 1983. Most of these countries will be conducting their data collection in mathematics classrooms at the 5th grade and/or 8th grade levels.

Research instruments to be used at the international level have been designed and are being translated and adapted to suit national requirements. They include a comprehensive set of classroom observation instruments to inform about the teaching context and classroom processes. Findings about what occurs in classrooms will be related to results on student cognitive tests and surveys of student perceptions of classroom processes. The training of observers makes up and important part of the study and workshops for this purpose have already been carried out in different world regions. Likewise, there is scope for the development of optional research instruments to be used in some countries only. Data collected at the international level will be analysed at a centralized data-processing location.

OUTCOMES.

Beyond its value for each country, it is expected that the study will produce knowledge of the kind available only from research conducted across national boundaries. In each country the study will have the same basic design, and will measure similar variables concerned with teaching practices and types of educational outcomes. The grade levels and subject matter will be similar in each country. The study should demonstrate useful methodological procedures and also lead to the improvement of teaching in many countries.

*Reprinted from IEA-Newsletter, nr. 1, january 1982.

Student activities and learning outcomes.

W. Tomic & E. Warries
Twente University of Technology
Department of Education
The Netherlands

INTRODUCTION

In this preliminary paper we shall report on results of a small part of the so called correlational study. The results in this paper are based on systematic observation of mainly student activities during mathematics lessons.
During the schoolyear 1981-1982 eight lessons of a sample of 50 mathematics teachers have been observed. Thus 400 lessons in the 8th grade have been recorded.
The study had two major aims:

1. To observe and measure the occurance of preselected teaching and learning activities in the classroom. In this paper we are focussed on student activities, the involvement of the student and the nature of the teacher's involvement in those activities.
2. To explore whether there exists a relationship between teaching and learning activities on the one hand and student cognitive and affective outcomes on the other hand.

The ultimate end is to translate the knowledge about these relationships in recommanded teaching practices. We want to draft a profile of an effective maths teacher.
In this paper we report on actual information about the course of things in the classroom which have been recorded rather objectively and precisely. Valid and fairly precise statements about e.g. student engagement during mathematics lessons are possible.

Way of reporting

The data which have been collected by the Classroom Snapshot Instrument in combination with student outcomes yield information to answer two questions:

1. How often do preselected student activities occur?
2. Which observed student activities are associated with learning outcomes?

To answer the first question the frequency scores and their standard deviations are reported. To be able to compare the student activities with one another percentages are given at the same time.
As to the second question two statistics are reported which reflect the relationship between the proces and the dependent variable: Pearson - and the

partial correlation coefficient. For reasons of interpretation we inspected scattergrams to explore the linear relationships between student activities and student outcomes.

Outline of the paper

In section 2 below we will briefly mention the observation system which has been used in the correlational study.
As the data presented in this paper have been gathered by the Classroom Snapshot Instrument we shall discuss this instrument in some detail. What follows is a description of categories of student and derived student activities. Then the role of the student and the nature of the teacher's involvement will come up for discussion. The focus then will be on the results with regard to frequencies and percentages of student activities and the role of the teacher. Next will follow an analysis of the relationships between student activities on the one hand and learning outcomes on the other hand.

THE OBSERVATION INSTRUMENT

The observation instrument that is used in the Dutch part of the CES was derived in large part from that developed by Jane A. Stallings at SRI International. This observation system is used to record student activities in the classroom and interactions between teachers and students.
The system as used in the Netherlands contains mainly two sections:
1. The Classroom Snapshot Instrument and
2. The Five Minute Interaction Instrument.
The last instrument registers in detail the interaction between the students and the teacher during the five minute observation periods of the class. In this paper we restrict ourselves to the results obtained by the Classroom Snapshot Instrument, a very small part of the so called correlational study.

The Classroom Snapshot Instrument

This instrument is used to indicate:
1. The diversity of activities during the observed mathematics lessons;
2. The number of students actively engaged in the various activities;
3. The number of students not engaged in the activities and finally
4. The role or the nature of the involvement of the teacher with his students.

The various activities recorded by the snapshot instrument were coded five times during one observed period of 45 minutes. So we have got a picture of the classroom activity on five separate moments.
The emphasis is primarily on student activities and secondly on the teacher's role.
Below we shall first give a description of the ten student activities, further the engagement of the student and finally the teacher's role.

Description of student activities

1. Listening to a lecture/explanation/demonstration.
   The students are listening to the teacher who is presenting academic information in the form of a lecture, explanation or demonstration. All kinds of materials can be used.
2. Reviewing previous work.
   The students are reviewing previous work, e.g. checking tests and assignments. This activity is directed by the teacher.

3. Participating in discourse/discussion.
   The students are interacting with the teacher e.g. the students respond to teacher questions which may form part of the general lecturing mode or an evaluative mode.
4. Participating in oral practive/drill.
   The students are participating in an oral practice or drill activity which does not form part of any evaluation.
5. Seatwork: taking tests.
   The students are taking a test or performing some formal evaluative task.
6. Seatwork: reading silently.
   The students are reading silently some subject-related books.
7. Seatwork: written assignments.
   The students are working on written mathematics assignments.
8. Seatwork: laboratory/manipulative.
   The students are working with laboratory equipment.
9. Non-academic.
   There is no academic activity in the classroom, Activities are transition, procedural, seating arrangements and disciplinary.
10. Other.
    This is a category for academic activities which does not fit in the above categories.

Description of derived student activities.

1. Variation in student activities.
   Number of different student activities that occur in 40 snapshots except non-academic activities. Maximum is 9.
2. Variation in student activities within one lesson period.
   Number of different student activities in a lesson period (five snapshots).
3. Variation in student activities at the same moment.
   The mean number of different student activities in one snapshot.
4. Amount of seatwork.
   Number between 0 en 1 indicating how often seatwork accurs followed by at least one other student activity. Seatwork at the end of a lesson is not taken into account.
5. Student participation.
   Number of students actively involved in the assigned academic activity devided by the total number of students, multiplied by 100: Percentage engaged students.

The role of the student.

1. Engaged students.
   These students are actively involved in the teacher assigned academic activity. This provides a measure of academic engaged time. It is not possible in this Classroom Snapshot Instrument to be engaged in activities which are not academic in nature.
2. Non-engaged students.
   If students are not academically engaged in the assigned task they are non-engaged.

Teacher's role.

1. Teacher is interacting with student.
   This means that the teacher is actively leading the group or interacting with one or more students.
   This role may be observed with student activities 1, 2, 3 and 4.

2. Teacher is monitoring.
   The teacher is monitoring or observing the students on an individual or a group basis. This role will often be observed with student activities 5, 6, 7, and 8.
3. Teacher is uninvolved.
   The teacher is not involved with the students. He may be working on administrative tasks and not monitoring student activities.

Student instruments

Of course the choice of instruments for measuring student cognitive and affective outcomes is important for this study. The initial and the final student questionnaire provided data on student background, attitudes and perceptions. The items used in the Netherlands for the greater part were identical to those used internationally. The variables for which data were collected by means of the student questionnaire were student characteristics such as sex, age, educational plans, the level of parental education, parental occupations and the language spoken at home.
The attitude variables included attitudes toward school, toward mathematics, self-related attitudes and sex-related attitudes. In the final student questionnaire items were repeated from the intial questionnaire on self-related and sex-related attitudes toward mathematics. Variables like perceptions of the classroom task-orientation and perception of both classroom instructional events and practices and of management events and practices were included.

As for the cognitive pretest we decided to use items already developed by the CITO. Ultimately 20 items were selected. For a rationale of items and for an extensive report on this subject, we refer to Krammer, 1982.
The cognitive posttest was mainly developed by ourselves and included 24 to 48 items. The reliability (KR 20) of the pretest was .61 and of the posttest .71.

RESULTS

Student activities.

As stated before in 'The Classroom Snapshot Instrument', in describing the Classroom Snapshot Instrument, the nature of the student's activity is first recorded. An overview of the results is given in Table 1.
Next to the mean frequencies also the mean percentages of the activities are reported, in order to make possible a comparison between the two.
As explained before non-academic student activities also were coded. These activities occurred in 15% of the cases. At first sight this seems to be a considerable loss of time on task. Of the preselected student activities 'reviewing previous work' occurs most. Over 27% of time students are engaged in previous work of subject matter which has already been dealt with. Reviewing previous evaluative tests comes under this heading too. After 'reviewing', the student activity which we have named 'listening' for shortness sake, comes next. In 23% of the observed cases students do listen to the teacher who is lecturing, explaining or demonstrating.
As for written assignments a mean percentage of about 22 has been found. Usually this means that students are doing mathematic problems/exercises. The mean percentage for participating in discourse or discussion is 11%. This means that the teacher within the scope of instruction or evaluation asks questions and the students answer his/her questions.
The remaining student activities like participating in oral practice or drill (we do not mean hearing lessons), making tests or doing other assignments in a formal evaluative situation, reading in a book and lastly seatwork with laboratory equipment, don't occur very often. With the exeption of the test this is actually understandable in mathematics education.

The teachers were asked not to give evaluative tests which will last longer then 15 minutes to their students during the observed lessons. This explains the low frequency of the student's activity 'test'.

Table 1: Classroom Snapshot Instrument: Ten student activities.

number variable	name variable	mean frequency	stan.dev.	mean %
M1	Listening	10.14	5.3	23
M2	Reviewing previous work	11.74	5.0	27
M3	Participating in discourse/discussion	4.86	5.2	11
M4	Participating in oral practice/drill	0.04	0.3	0
M5	Seatwork: test	0.40	1.0	1
M6	Seatwork: reading	0.02	0.1	0
M7	Seatwork: written assignments	9.40	6.2	22
M8	Seatwork: laboratory	0.04	0.3	0
M9	Non-academic	6.52	2.6	15
M10	Other	0.14	0.3	0
				100

Relationships between student activities and learning outcomes.

In Table 2 correlation coefficients between student activities and learning outcomes are reported. The five most frequently occuring activities were listening, reviewing previous work, participating in discourse/discussion, written assignments and non-academic activities (see Table 1). Two of these activities are negatively associated with student achievement, namely written assignment and non-academic student activities. This means that working independently on mathematic problems is negatively correlated with learning outcomes. Obviously the frequency of this alterable student activity should be diminshed by the teachers. Written assignments however take up an important part of the lesson: a mean percentage of 22 has been found. Non-academic activities also correlate negatively with achievement, i.e. in general the more time spent on non-academic activities, the worse the achievement. This sounds plausible indeed. Further there is a trend for a positive association between reviewing previous work and both cognitive and affective outcomes. This activity occurs rather frequently (mean 27%) in classroom practice.
Two relative high correlation coefficients have been found for the variables 'test' and 'reading'. In the table 2 is shown that these results are little realisitc: there are no linear relationships. At last it is notable that the student activity 'listening' is positively associated with student attitudes. One possible conclusion is that students like this - for themselves - rather passive occupation.

Table 2. Correlations between student activities and learning outcomes.

number variable	name variable	correlation with cognitive posttest	partial correlation with cognitive posttest	partial correlation with affective posttest
M1	Listening	.04	.04	.24**
M2	Reviewing previous work	.20*	.19*	.20*
M3	Participating in discourse/discussion	.02	.02	-.25**
M4	Participating in oral practice/drill	.10	.09	-.05
M5	Seatwork: test	-.28[1)]	-.27**	-.44***
M6	Seatwork: reading	-.28[1)]	-.27**	.05
M7	Seatwork: written assignments	-.18	-.18	-.10
M8	Seatwork: laboratory	-.15	-.15	-.10
M9	Non-academic	-.18	-.19	-.03
M10	Other	.06	.05	.09

* p = .10
** p = .05
*** p = .001
1) = non linear relationship according to scattergrams.

Derived student activities.

In this section we shall pay attention to five so called composite variables derived from the directly observed singular variables. A summary of these variables including mean scores and standard deviations is given in Table 3. Each derived variable is followed by an explanation so that the description of the derived student activity gets clear. The calculation of the mean scores in question is explained too. It is obvious that the reported mean scores cannot be compared with one another. From the data it appears that the average number of student activities (SS1) is five. By way of explanation we mention that the total number of observed student activities in the Classroom Snapshot Instrument is nine. So the maximum value of derived student activity SS1 equals nine. In the section 'Student activities' we already mentioned which activities occur most.
As far as variation in student activities within one lesson, a mean score of 0.44 has been found. Because five snapshots per lesson were recorded, this means that on an average at least two different student activities occur per one lesson. By means of the Classroom Snapshot Instrument at the same time it was recorded whether during one snapshot different student activities were going on (SS3). Looking at the mean score in question (1.07) this does not appear to be the case.

So there is no evidence for variation in student activities at the same moment. Explored as well is how often self-activity by students occurs (SS4). We don't mean seatwork at the end of the lesson, but a knowingly selected and planned activity amidst the other observed student activities. Seatwork defined in this manner still occurs to a considerable extent. Finally and that is interesting, by combining singular variables from the Classroom Snapshot Instrument, we can get an indication of student participation. The proportion of actively involved students in the assigned academic activity is over 73%. At this stage it is not advisable to give our opinion about the acceptability of this finding.

Table 3. Classroom snapshot instrument: Five derived student activities.

Number variable	Name variable	Explanation	Mean frequency	Stan. dev.
SS1	Variation in student activities	Number of different student activities in 40 snapshots (8 observed lessons) except non-academic. Maximum: 9	5.00	0.53
SS2	Variation in student activities within one lesson period	Number of different student activities per lesson period (5 snapshots)	0.44	0.05
SS3	Variation in student activities at same moment	Mean number of different student activities during one snapshot	1.07	0.07
SS4	Amount of seatwork	Number between 0 and 1 indicating how often seatwork occurs, followed by at least one other activity	0.14	0.15
SS5	Student participation	Number of students actively involved in the assigned academic activity devided by the total number of students multiplied by 100: percentage engaged students	73.29	10.21

Relationships between derived student activities and learning outcomes.

In table 4 there are three process variables with reference to variation in student activities (SS1 to SS3). Variable SS1 refers to the total number of different activities which occurred in the eight observed lessons per teacher. According to the scattergrams there is no linear relationship between this derived variable and student achievement. The partial correlation coefficient with student attitude is negative; however no data are available about the linearity of the relationship. Neither the occurrance of different activities within one lesson (SS2) nor the occurrance of various activities at the same moment (SS3) are associated with student achievement. For achievement it is of no consequence whether the teacher varies frequently student activities.
Variable SS4, amount of seatwork amidst, is calculated to allow for this derived activity to be distinguished from self-activity at the end of the lesson. For as we know the last mentioned activity is frequently practised by teachers to fill up the remaining time, so that the students can start with their homework meanwhile. With the derived variable self-activity during the lesson we tried to measure to what extent self-activity is selected knowingly by the teacher. Practices which demand self-activity from the students are: reading, written assignments and laboratory-work.
There is no association with achievement. With regard to the three single activities (M6, M7, M8) there is a negative relationship with achievement, see table 2. Possibly these activities have a negative effect on achievement in so far as they are applied at the end of the lesson. The derived variable student participation (SS5) correlates positively with the criterium variable cognitive posttest. That is to say the greater the proportion of students actively involved, the better the achievement. This is in conformity with former research by others in the corresponding concept 'academic learning time'.

Table 4. Correlations between derived student activities and learning outcomes.

Number variable	Name variable	Correlation with cognitive posttest	Partial correlation with cognitive posttest	Partial correlation with affective posttest
SS1	Variation in student activities	-.14	-.13	-.35
SS2	Variation in student activities within one lesson period	.07	.05	-.02
SS3	Variation in student activities at same moment	-.15	-.13	-.14
SS4	Amount of seatwork	-.05	-.03	-.10
SS5	Student participation	.31*	.32*	.06

*p = .01

Teacher's role.

As mentioned before, the nature of the teacher's involvement was considered as well: teacher is interacting, is monitoring or is uninvolved in student activities. A summary of the three observed role's and their explanation are to be found in table 5. At the moment of this report standard deviations were not available, hence they are missing in this table.
In more than 78% that was recorded in the classrooms by means of the Snapshot Instrument, the teacher was interacting with the students. Over 18% of the observations he/she was monitoring. Only in 3% of the observed lessons the teacher was not involved in students' activities. The nature of the teacher's involvement consists by far for the greater part of interacting with students.

Table 5. Classroom snapshot instrument: Nature of teacher involvement.

Number variable	Name variable	Mean percentage	Explanation
SS6	Teacher interacting	78	Teacher is actively leading the group in the activity which is on or is interacting with one or more students.
SS7	Teacher monitoring	19	Teacher is monitoring or observing the students on an individual or a group basis.
SS8	Teacher uninvolved	3	Teacher is not involved with the group, e.g. he may be working at his desk and marking papers and not monitoring the students while they were doing seatwork.

Concluding remarks.

1. As mentioned before in the first section both product-moment correlation coefficients and partial correlation coefficients are calculated. The main reason for partialling out cognitive pretest scores from cognitive posttest scores is the fact that cognitive entering behavior is assumed to explain much variance in the student posttest scores. By partialling out pretest scores it was possible to explore to what extent there remains a relationship between student activities and product variables after correcting the influence of the variable student cognitive pretest.
   When we compare the correlation coefficients which have been calculated in different ways, we can observe that there are only slight differences.
2. The use of derived student activities measures was meaningful in our opinion. In this way we have obtained more, and more detailed, knowledge about alterable student activities in the observed classes.
3. With this relative simple observation technique it appears to be possible to record interesting variables. The technique could also prove useful for descriptions of lessons in other school subjects.
4. At least two recommandations for teachers seem indicated from the data:
   - try to get as much attention from your class as possible when you are interacting.
   - Do not hasten to finalize your teaching before the end of the period through seatwork.

REFERENCES

Berliner, D.C.: Tempus Educare. In: P.L. Peterson and H.J. Walberg (Eds.) *Research in teaching: Concepts, findings and implications.* Berkeley, California: McCutchom Pub. Co., 1979.

Fisher, C.W., Filby, N.A., Marliave, R.S., Cohen, L.S., Dishaw, M.M., Moore, J.E., Berliner, D.C.: *Teaching behaviors, academic learning time and student achievement: Final report of phase III-B, Beginning Teacher Evaluation Study (Technical Report V-1).* San Francisco: Far West Laboratory for Educational Research and Development, 1978.

Rosenshine, B.: Content, time and direct instruction. In: P.L. Peterson and H.J. Walberg (Eds.) *Research in teaching: Concepts, findings and implications.* Berkeley, California: McCutchom Pub. Co., 1979.

Stallings, J. and Kaskowitz, D.: *Follow through classroom observation evaluation, 1972-1973.* Menlo Park, California: Stanford Research Institute, 1974.

Stallings, J.: *Learning to look: A handbook on classroom observation and teaching models.* Belmont, California: Wadsworth, 1977.

Tomic, W., Weesie, C.A.M.: *Leerlingactiviteiten, lesgedrag van leraren en leerresultaten.* Enschede: T.H. Twente, Toegepaste Onderwijskunde, 1983.

Section V

The Second IEA Science Study.

General Overview.

J.P. Keeves
Australian Council for Educational Research
Melbourne
Australia

The late 1950s and early 1960s were associated with a view of curriculum development in science and mathematics, that in a relatively short period of time changed science education in both developed and developing countries. This was the context with which the First IEA Science Study was planned and conducted in 1970. Nineteen countries took part in this study, and both the published reports and the data and documents held in archives provide an unequalled synoptic view of science education at that time. By the early 1980s this wave of development in science education had come to an end. The Education Division of the National Science Foundation in the United States was to be closed. The Schools Council in England and the Curriculum Development Centre in Australia, bodies which had taken over the initiatives for new work in the area of science education, were to be terminated. At least in English-spoken countries this was the end of an era, and quite clearly a point in time at which it was essential that a further detailed examination of science education should occur.

However, nothing remains stationary - all is in change. The advent of the very powerful micro-computer and other new technologies have within two or three years, brought a renewed interest in science education. Three important questions are being asked of science education. First, 'what is the contribution of science education to the developments in micro-electronics, information technology and bio-technology?' Secondly, 'what contribution can micro-electronics and the new technologies make to the teaching of science?', and thirdly, 'can Science Education provide a sound knowledge and understanding of the environmental impact of the new technologies on ourselves and the world in which we live?'. A new wave of curriculum development has not yet started. There is, however, a critical examination being undertaken in many countries of 'what science is being taught?' 'what science should be taught?' and 'how should science be taught?' The challenge to those of us engaged in the Second IEA Science Study is that of making a major contribution to this debate.

In this context it has been essential during the planning of the study that we should maintain an appropriate balance between the needs of an international comparative study that will allow cross-national comparisons to be made, and the needs of the national studies that will examine critically the issues for science education in particular countries. We are hopeful that up to 30 countries will be taking part in the international study during the years 1983

and 1984, and that each country will not only contribute effectively to the international data base, but will also obtain the information necessary for a full consideration of the questions that are relevant for the future planning of science in that country.

The five basic aims of the study are:

1. to measure, by means of large-scale survey procedures, the current state of science education in schools across the world;
2. to examine the ways in which science education has changed since 1970;
3. to identify the fsctors which explain differences in the yields of science education programs across countries, and between students within countries, with particular attention to the role of the science curriculum as an explanatory factor;
4. to investigate changes in the patterns of relationships between the explanatory factors and the yields between 1970 and 1983; and
5. to assist all participating countries, especially the less developed countries, to carry out national studies of science education in order to investigate issues of particular interest in their own countries.

It is an immense task that we have undertaken. Nevertheless it is, we believe, an important task for IEA, and one that is consistent with the role envisaged for IEA to conduct research studies that are comparative, cooperative and universal and which will contribute to the endless quest of building a body of knowledge and understanding about education across the world.

Optimalization of reporting results from national Assesment Studies.

W.J. Pelgrum
Twente University of Technology
Department of Education
The Netherlands

INTRODUCTION

Since March 1983 The Netherlands is participating in the Second International Science Study (SISS) of IEA.

In this study, just like some other studies of IEA (such as the Second International Mathematics Study) a description is made of the content and outcomes of education in a certain subject matter (in this cas science) at certain gradelevels of all sectors of a national educational system. Although the Ducth participation in this study is restricted to the third grade of secondary education, the study can be directed through the addition of national options to a number of questions which are of special interest in the Dutch situation. At this moment a number of subject matter oriented questions (which have been asked by curriculum-developers, teacher educators, school inspectors) are being worked out.

The Dutch participation in SISS is not only interesting for reasons of subject matter, but also because in our country there is relatively little experience with this type of research which is strongly related to assessment studies which are performed in other countries (United States for example NAEP, CAEP) or England (APU). Recently the Dutch Ministery for Education and Science started a study in which the feasability of national assessment in the Netherlands in primary schools is investigated. National assessment is a type of research from which instruments and data become available from which at different political level optimalization-measures can be derived.

In the Netherlands this type of research has been discussed by some authors (Wijnstra, 1982; van der Linden & Pelgrum, 1983). It was shown that national assessment aomong other things raises problems as far as the reporting of data is concerned. In this paper we will discuss this problem and offer a conceptual framwork in which this problem can be located. In the end a number of research questions will be offered which can act as a starting point for the elaboration of a Dutch option in SISS.

FUNCTION OF NATIONAL ASSESSMENT

National assessment is primarely for the use of educational policy-making (at several levels). It is embedded in a cyclic proces of quality control in which the next four stages can be distinguished:

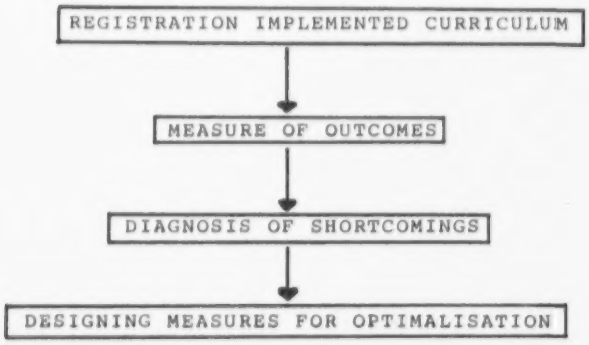
1. Identification of standards.
2. Perception of the degree of standard realisation.
3. Evaluation of stage 2 data.
4. Construction of measures (to maintain or alter standards).

The educational researcher designs techniques, performes necessary background analyses and produces the data which are needed in stage 1-3. From the differences between standards and observed score-profiles measures for optimalisation can be constructed, after which the cycle is repeated in order to study the changes which take place over time.

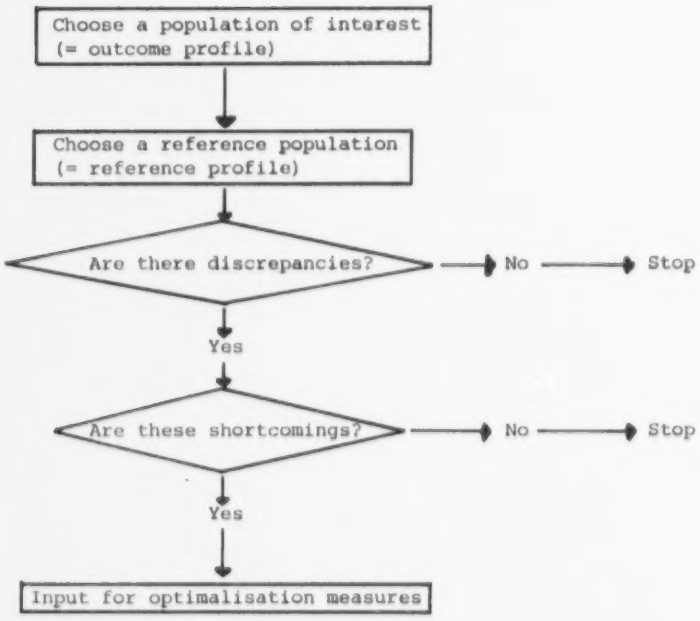
The stage of standard identification is of great importance. It is not an easy task to identify uniform standards. Especially in a relatively decentralised educational system, like the Dutch, uniform standards can - generally spoken - hardly be identified (except at a very global level). An alternative for absolute standards is to work with relative standards, whereby observed score-profiles of (sub)populations of students are compared with score-profiles of other (sub)populations in order to gain insight into the question if improvements could be made. Note that the use of the term "relative" here is not identical with the relative procedures in psychometrics for the determination of cutting scores. Later we will return to this distinction.

The aforementioned problem of identification of standards in relatively decentralised educational systems is caused by the absence of uniformly operationalised curriculum prescriptions. Of course, operationalisations in the form of final exam prescriptions are present, but these are only a limited reflection of the goals which are pursued (for restricted groups of students), let alone that they reflect outcomes of education.

For this reason it is useful to generalise the above-mentioned assessment cycle according to the following conceptual model:



The stage of diagnosis in this model can globally be described as follows:



The advantages of this way of conceiving an assessment cycle are as follows:
- The implemented curriculum (and especially the variation therein) is an explicit component of the cycle.
- Absolute standards are not strictly required (but can be incorporated by taking a reference score-profile which is derived from standards).
- Measures can be derived from discrepancies between the intended curriculum (outside the model) and implemented curriculum as well as from discrepancies between observed score profiles and reference profiles.
- The model can be applied at different educational political levels (by aggregating data at different levels and choosing different reference profiles. At a national level reference profiles can be derived from other countries).
- The measuring of outcomes can be directed at the union of implemented curricula such that exceptional curriculum operationalisations are not excluded.

EXECUTION OF NATIONAL ASSESSMENT

In the execution stage of national assessment the following sub-stages can be distinguished:
1. Identification of the curriculum-domain
   (in sub-domains).
2. Sampling of curricular elements
3. Construction of instruments
   (including pilot-testing and modifying).
4. Data-collection
   (including background questionnaires and registration of the implemented curriculum).
5. Calculation of generalisable domain-scores.

(In case of the use of item-banks stages 1-3 can be reduced to sampling items from the bank).
In national assessment an adequate coverage of the total subject matter requires the use of large item samples. The testing time for students can however be limited, due to the technique of multiple matrix sampling, whereby samples of items are presented to samples of students.

For the calculation of testscores or subtest-scores several psychometric techniques are available. Later we will go into some of these techniques and into the problems of registering the implemented curriculum.

REPORTING NATIONAL ASSESSMENT DATA

The reporting of data from national assessment has as its main goal (such as indicated in the aforementioned model) to diagnose shortcomings by several educational agents at different levels (national as well as local). Besides, we have to assume that not always uniform standards are present, such that one needs to work with relevant reference-groups.

In order to realise the main goal of reporting the question is important how the diagnosis of shortcomings by specific usergroups could take place. As indicated in the model at page 3 the identification of discrepancies does not directly result in the identification of shortcomings. For this an additional interpretation-step is necessary. An example might clarify this. Suppose that in a study a biology test is used which covers adequately the following areas:

1. Cell structure and function.
2. Transport of cellular material.
3. Cell metabolism.
4. Cell responses.
5. Concept of the gene.
6. Diversity of live.
7. Metabolism of the organism.
8. Regulation of the organism.
9. Coordination and behavior of the organism.
10. Reproduction and development of plants.
11. Reproduction and development of animals.
12. Human biology.

(adapted from the content-grid SISS)
Suppose in addition that subtest-scores (score) related to these areas and the degree of curriculum implementation (Imp) are known and that the results of different subgroups of students compared to reference-groups can be considered.

In figure 1 a fictitious example of some possible comparisons is presented.

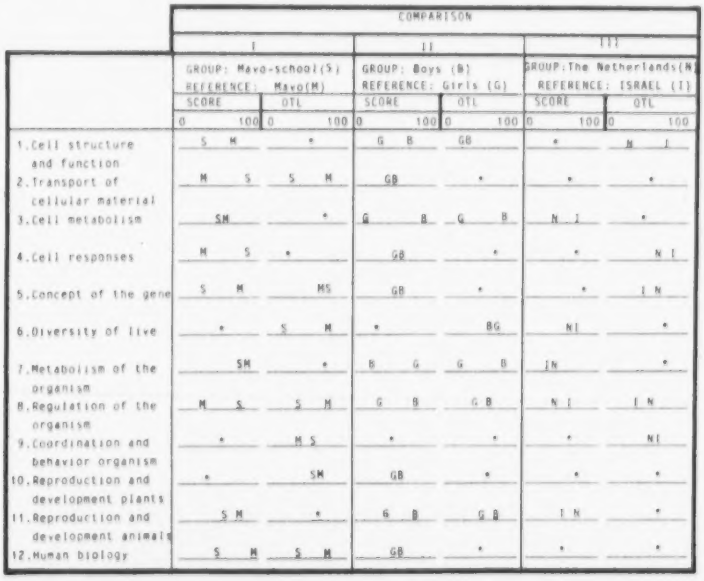


Figure 1: Example of assessment results.

The perception of these score profiles does not directly result in the identification of shortcomings. In comparison I potential areas are: 1, 5 and 12. In area 5 probably specific problems exist considering the relatively high degree of implementation compared to the relative low score. The results in area 12 are relatively bad, but this might be caused by a relative low degree of implementation. Comparison 2 shows that girls receive less biology training than boys. Comparison 3 shows that in the Netherlands possible shortcomings exist in the areas 3 and 8, even in spite of the high degree of implementation in area 8.

This example shows in a simple way that there are techniques to identify shortcomings, provided that the right referencegroups and the relevant context information is used. In the example the degree of curriculum implementation was chosen as the relevant context-information, but it is also possible - when needed - to compare groups on other contextual variables. In this example the user determines which information is needed for which goal.

Very little is know about the informational needs of potential users of assessment-data-banks as well as the goals for which the information has to be used.

One can easily assume that in the educational field a big variation in informational needs exists. For the sake of reporting this variation raises problems, because for practical reasons it is almost impossible to incorporate in one report all theoretical interesting comparisons between profiles. The introduction of micro-computers in schools can in this respect increase the possibilities, because by the way of fexible procedures, information extraction by the user can take place.

A problem which has not been mentioned yet is that due to the variation in implementation the calculation of groupscores on tests cannot take place directly. One has to avoid the problem of incomparability in trying to compare groups with implementation related to different subtests of items. As a consequence artificial discrepancies (or non-discrepancies) might result caused by item characteristics.

A last problem is the quantification of implementation. A good quantification is needed for a justified interpretation of data resulting from national assessment. Currently not much is known about the question which instruments are suitable for measuring the implemented curriculum in the context of national assessment.

The aforementioned problems will be discussed in more detail in the next section.

PROBLEMS AND POSSIBLE SOLUTIONS

In the preceeding three main problem-areas associated with the use of data from national assessment were identified. These problems are:

1. Informational needs of users.
2. Testscore calculation in heterogeneous curriculum settings.
3. Quantifications of curriculum-implementation.

In the next section a first exploration of these problem areas and possible solutions will be presented.

Informational needs of users.

At this moment - to the author - not much is known about the informational needs of potential users of data from national assessment. However, experiences abroad show that more should be known. For example: one of the conclusions of the American General Accounting Office in 1976 after critically evaluating NAEP was: (see Wijnstra, 1982, page 14)

".... redirect the project by identifying the informational and other needs of decisionmakers ...."

It is not known if this recommendation really resulted in an assessment of informational needs.

Scheerens (1983) concludes in another context that:

"In the literature on evaluation and policy oriented research rightly a growing attention is paid to the use of research results by policy-makers".
And later on:

"In the first place it is striking that "use" in different empirical studies is differently operationalised, amongst other as "utility" (such as judged by researchers and policymakers), awareness of results, being influenced by certain concepts and reformulating problems and direct "tangable" use in the form of concrete decisions of policy modifications".

In the first place a more profound study of this research literature has to take place in order to investigate to what degree the conclusions can be generalised to the use of data from national assessment. One might expect however that the studies are focussed on the use of data and results from written reports.

Therefore, in connection with the goal of constructing user tailored reports additional research concerning the informational needs of potential users will be necessary. The main components of such a research-project would be:

I Construction of a catalogue of variables

II Identification of potential users

III Specification by users of desired information

IV Analysis of informational needs

V Construction of user tailored reports

VI Registration of actual use and opinions on utility

Although the feasability of this study in a limited setting is possible, an important issue for investigation will be the relevance and feasability of the implementation of the reporting procedure on a large scale.

TESTSCORE-CALCULATION RELATED TO VARIATION IN IMPLEMENTED CURRICULA

In the preceding paragraphs the variation of implementation of curricula of certain schoolsubjects in decentralised educational systems was mentioned.

A consequence is that in the case of national assessment the test-item collection will not cover to the same degree the actual implemented curricula in all sections of an educational system. This poses no problems for the estimation of domainscores. In that case the compentences of students in the total subject area have to be estimated. Problems occur however in comparing profiles: in that case it is not clear to what degree discrepancies between profiles can be attributed to item-characteristics: it is possible that the implemented curricula behind the profiles were oriented to items with different complexity. The question is to what degree the measures are equivalent. In this case we are dealing with the problem of test-dependent scoring. Solutions for this problem can be found in applying item-respons models (van der Linden, 1978), whereby population-independent estimates of item-complexity can be made. A problem however is that these models don't allow extreme deviations of item complexity in different subpopulations. Therefore it is necessary to investigate to what degree it is theoretically and practically possible to perform the item calibrations on curricular homogeneous subpopulations of students. An adequate solution to this problem contributes to an interpretable representation of data from national assessment in case profile comparisons are made at different levels.

QUANTIFYING THE IMPLEMENTED CURRICULUM

Data on the implemented curriculum can be collected at different levels of specificity. At a global level the actual time expenditure for a total school subject or subtopic of it are interesting. The Second Mathematics Study showed

that in the Netherlands in grade level 8 a substantial variation in the actual time expenditure in mathematics exists. Besides global information however it is als useful to collect data on a more concrete level.

While interpreting the data from national assessment it is important to know how much and which test-items belong to the actual implemented curriculum. On the one hand this contributes to a better understanding of the degree of coverage of the curriculum by the test-instrument and at the other side the possibility is created to take account of these differences when performing profile comparisons.

One of the instruments with which experience has been gained in several IEA studies is the so called Opportunity to Learn questionnaire. This questionnaire has been revised a number of times. In the most recent versions this questionnaire traces when the subject matter of which the test-item are operationalisations will be or has been offered to the students. Teachers make this judgement for each test-item apart.

In the international report on the first Science Study tables are presented which show besides testscores also an associated Opportunity to Learn index. From these tables it can be calculated that (in the population of 14 year old students) the correlation between testscores and Opportunity to Learn for 17 at country-level aggregated cases is .73 (Comber & Keeves, 1973, table 7.2. population II).

Regression-analysis within countries showed however that the contribution of the Opportunity to Learn variable is very low. The interpretation of this finding is however hampered because the interdependency of (blocks of) variables. As a consequence the sequence of introducing variables in the analysis is important: an earlier introduced variable type of school "explains" a lot variation while for opportunity to learn little variations remains to be "explained". Altering the sequence would probably have had the opposite effect. The Second Mathematics Study also used an Opportunity to Learn questionnaire. Pelgrum, Eggen & Plomp (1983) present the results of some analyses on data which were collected with this instrument in the Netherlands. The authors conclude that there are indications that this instrument is valid for the identification of the implemented curriculum in the first two years of secondary education. Judgements by secondary school teachers of actual contents of the curriculum in elementary school conflict however with the information form an other reliable source. According to the authors ongoing analyses are necessary to gain a better understanding of the quality of this instrument.

CONCLUDING REMARKS

The preceding paragraphs dealt with one of the problems associated with national assessment, which qua method resembles SISS. The main goal of SISS is the description of content and outcomes of education of one school subject in certain grade levels of all sections of a national educational system. For an optimal use of data from this study special reporting procedures have to be developed, taking account of the diversity of user-groups and the variation of actual implemented curricula. For this a better understanding of the nature of informational needs of pontential users of the data-bank is necessary.

For the use of interpretation an adequate registration of the actual implemented curriculum is needed. In the preceding paragraphs it was indicated how research can be directed to these problems. In this research the problem of incomparability of testscores of sub-populations caused by the variation of actual implemented curricula needs to be resolved, such that in spite of this variation comparisons of score-profiles are possible.

We intend to aim one of the Dutch options in SISS on the aformentioned cluster of problems. The final goal of this enterprise is to contribute to the design of directives and procedures which optimise the use of data from national assessment or in the words of Wirtz and Lapointe (Lapointe, Koffler, 1982):

"The assessment program should be designed and administered to optimize it's service function to state and local edcuational assessment and standard setting agencies".

REFERENCES

Comber, L.C. & Keeves, J.P.: *Science education in nineteen countries.* Stockholm, International Association for the Evaluation of Educational Achievement (IEA), 1973.

Lapointe, A.E. & Koffler, S.L.: Your standards or mine? *Educational Researcher,* Dec. 1982, p. 4-9.

Linden, W.J. van der: Het klassieke testmodel, latente trek modellen en evaluatie onderzoek. *VOR publicatie nr. 7,* Amsterdam, Vereniging voor Onderwijsresearch, 1978.

Linden, W.J. van der & Pelgrum, W.J.: *Periodiek onderzoek naar de kwaliteit van onderwijs: Enige practische en methodologische problemen.* Enschede, T.H. Twente, Onderafdeling Toegepaste Onderwijskunde, 1982.

Pelgrum, W.J., Eggen Th.J.H.M. & Plomp, Tj.: *Some results of the Second International Mathematics Study in the Netherlands.* Paper for the 24th IEA General Assembly, Enschede, August 1983.

Scheerens, J.: *Evaluatie-onderzoek en beleid.* s'Gravenhage, Stichting voor Onderzoek van het Onderwijs, 1983.

Wijnstra, J.M.: Wat leren kinderen op school? *Pedagogische Studiën,* 1982, vol. 59, p. 223-232, p. 277-287.

Comment on the Dutch paper.

J.P. Keeves
Australian Council for Educational Research
Melbourne
Australia

Hans Pelgrum is correct in the emphasis in his comments that there is concern in many of the countries taking part in the Second IEA Science Study for the improvement of the quality and relevance of science education within the particular country. To enable judgments to be made that have some meaning, it is necessary to identify standards and make assessments of the degree to which the standards have been realized. However, both in the identification of standards and in the assessment of the extent to which standards have been realized, strength and validity can be obtained by the making of comparisons both across countries and across time. With these purposes in mind, in the planning of the study great care has been taken to ensure that meaningful comparisons can be made with the First IEA Science Study and to obtain consensus between the participants in the study that the tests used will assess with validity the different science curricula of the countries taking part.

In addition, Hans Pelgrum has also emphasized that the study should be designed so as to optimize its contribution to the investigation of problems in science education at both the national and more regional levels. The study has not been designed to provide data of consequence to individual students, teachers or schools. Indeed it is important that reassurance should be given to those taking part that individual students, teachers and schools will not be identified. However, the examination of the data which will be undertaken and reported will be carried out at the levels of analysis of between students, between schools or between classrooms, and between countries. In this way general statements will be made that relate to the factors affecting yield in science education for students, schools and nations. Some countries, have, in addition, designed their samples so that information will be available for recognizable regional units within the country. In Australia, the six States and two Territories will be considered, while in Canada, the country has been subdivided not into provinces but into four major zones which combine provincial regions. In so far as these sub-national units have common characteristics in their provision of science education, such a breakdown of the national data can be extremely valuable for the future development of science education programs within the country.

The First IEA Science Study produced, I believe, three very important findings. First, there were significant differences between the sexes in level of achievement in science. These differences were greater in Physics and

Chemistry and less in Biology. Moreover, while strong sex differences existed at the 10-year-age level, successively greater differences were recorded at the 14-year-old and terminal school levels, in spite of a fall in participation by girls in the study of science. Secondly, the most consistently powerful variables operating at the student, school and national levels were associated with the time spent by the students in learning science and the opportunity that the students had to learn the content being tested. These effects were successively greater from the 10-year-old to the 14-year-old to the terminal secondary school level. Thus there was convincing evidence that the work of the school and the science curriculum provided by the school were significant factors affecting learning. While such results are not surprising, they are important and reassuring for those involved in planning the science curriculum. Thirdly, wherever teaching and learning practices were identified as making a contribution to accounting for the differences between students and schools in level of performance in science, these factors appeared to be associated with systematic and planned teaching. It would seem that the manner in which the curriculum is taught does influence learning, and while the significant variables differed markedly both across age levels and across countries, the planned and purposeful implementation of the curriculum in science is of consequence. This evidence from the First IEA Science Study has indicated to us that in the second study we should be planning to investigate as thoroughly as possible, and in ways consistent with the first study, the science curriculum as it applies differently to male and female students and the manner in which it is taught or implemented.

The analysis of the Curriculum

In the First IEA Science Study we recognized that the curriculum was being examined at three levels: (1) the prescribed curriculum, or the intended curriculum as laid down in the authorized syllabuses used within a country and within schools, (2) the translated curriculum as assessed in terms of the opportunity that the students had to learn the content-tested, and (3) the achieved curriculum as measured by achievement on the tests that were employed. The most difficult of the three curriculum levels to examine was the prescribed curriculum. In countries where a national syllabus was laid down, a syllabus statement existed, but such a statement was difficult to compare with similar statements which were available in other countries. However, in countries where responsibility for the curriculum was devolved to schools or to individual teachers it was much more difficult to obtain a simple and a coherent view of the curriculum. In the Second IEA Science Study to tackle this problem and to make effective comparisons between teachers, schools and countries, each, as appropriate, has been asked to provide curriculum ratings on a four point scale on the same 57 basic content areas of the science curriculum that were used in the first study in 1969-70. This was necessary so that some comparisons across a 14-year period to 1983-84 could be made. In addition, information was sought on further areas including the History and Philosophy of Science, Environmental Science, Technical and Engineering Science, Rural Science, and Health Science. Furthermore, an attempt was made to obtain similar data on nine process areas concerned with the processes of scientific inquiry. Neither the additional curricular areas nor the processes of scientific inquiry were sufficiently emphasized in any country, and rarely in individual schools for the information recorded so far to be of use. However, the data obtained across countries has not only been of direct value in the tasks of test construction to ensure that a sound sampling of topics was carried out by the testing program, but it has also provided evidence of recognizable differences in emphasis in the sicence curriculum both within and between countries. We are now very hopeful that this information will be of use in reporting the findings of both national and international studies.

Likewise, we are seeking information on the opportunity that the students had to learn the content tested using similar rating procedures. Here we are on firmer ground, because strong relationships between opportunity to learn and performance on the achievement tests were reported from the Mathematics Study

in 1964, and extremely interesting relationships were obtained from the First IEA Science Study in 1970. These results reassure us that although the data from individual teachers and schools may contain some error, when these data are aggregated to a regional or national level, important and valuable relationships are obtained. We recognize that there are inherent dangers in attempting to quantify the science curriculum, but the numerical data that are thus made available enable far stronger analyses to be carried out. The major problem that we currently face if we obtain relationships of interest is one of how to present the pattern of relationships without obscuring detail or losing the strength of the relationship recorded.

In the Second IEA Science Study we are hopeful that we will be able to examine the different approaches to the science curriculum in different countries and in different schools. Our concern is to provide a sound basis, as Hans Pelgrum has pointed out, for an investigation into the strengths of the science curriculum in the countries engaged in the study and where appropriate to diagnose weaknesses. Hopefully, we will develop a tool that will also be effective enough to be useful subsequently in making comparisons between the science curricula of different schools, so that where responsibility for the curriculum is devolved to schools it will be possible to provide assistance with the tasks of curriculum planning and development.

The challenge to those of us involved in the planning of the Second IEA Science Study is to provide information on the prescribed, translated and the achieved curriculum that will not only be useful in those countries where the science curriculum is laid down centrally, but will also be useful to individual schools and teachers who are responsible for developing and implementing the curriculum at the school level.

Section VI

Schooling and Equality.

Schooling and Equality

J.S. Coleman
The University of Chicago
U.S.A.

The ideal of equal educational opportunity is one that has come to be increasingly widely held throughout the world. In highly developed nations and in less developed nations, the ideal is expressed often and with vigor. If there is any theme in education more dominant than any other in nations throughout the world, it is this theme of equal educational opportunity for all children within a nation. In every nation, there is a general recognition, in the government and in the populations, among educational professionals and amoung lay persons, that the ideal is far from being realized. Yet the demand for equality of opportunity is a strong and widely shared one.

It was not always so. In the early years of public education in Europe, there was no thought of equal educational opportunity. The educational system followed the pattern of the class structure, with a low-level common school attended for a brief period by children of commoners, and an elite tier for those from higher backgrounds and destined for higher occupations. And since schooling was a local community activity, each community determined its own level of educational effort, with no thought of equality between communities. In America, without the background of a feudal class structure, the ideal of the single common school for all was present from the beginning of public education; but the local community responsibility for and authority over education meant that there was never a conception that all children throughout the nation were entitled to an equal educational opportunity. It is still the case in America that legal requirements for equalizing educational opportunity are limited to within each of the fifty states. In less developed countries, where state-proveded public education is rather recent, the edcuational system seldom encompasses all children, and ideals of equal educational opportunity are even farther from realization.

Yet the ideal is an exceedingly strong and widely held one. Why is it that the ideal has gained such strength, in diverse countries throughout the world? What are the social conditions that have brought demands for equality into being? And given that the ideal is strong, just how does a nation's educational system go about providing equal opportunity in education?

As it will turn out, answers to these last two questions are related. The provision of something approaching equal educational opportunity differs in different circumstances. A nation with one kind of social and economic structure can approach equal educational opportunity in a very different way than can a nation with a different social and economic structure.

But is is first useful to be a little more explicit about what is meant by equal educational opportunity. This is a task I find particularly familiar, because I addressed much the same question in some detail almost twenty years ago. The time was the year following passage of the Civil Rights Act of 1964 by the U.S. Congress. In a section of that Act, the Commissioner of Education was requested by Congress to assess the 'lack of equality of educational opportunity by reason of race, religion, or national origin'. I was in charge of the project to do this, and our first task was to discover just what Congress had meant by the phrase 'equality of educational opportunity'.

As it turned out, there was no single meaning at all. The phrase meant one thing to one Congressman, another thing to a second. And there are a large number of Congressmen. In the end, we had to select out dominant classes of definitions, and provide information about the 'lack of equality' according to these definitions.

There were two dominant classes of definitions, one having to do with inputs into education, and the other having to do with outputs from the educational system. The first class of definitions was concerned with things like financing of education, age of textbooks, size of physical facilities, size of library, qualifications of teachers, and other tangible resources which go into schools. The second class of definitions was concerned with what the schools produced: Proportion of children in a given location on from a given group who finished high school, special vocational skills learned, and most frequently of all, scores on standardized tests.

Advocates of the first class of definitions argued that the role of the government in providing equal educational opportunity lay in providing equal access to educational resources by all children. Whatever outcomes this produced, all children had the same fair chance at the resources of which schooling consisted. This might produce quite unequal outcomes, for some children were better able to put these resources to use, to take advantage of them, than were others. These advocates argued that equality of educational outcomes could only come about through unequal educational opportunity, withholding educational resources from those very children most able to profit from the resources. A secondary impact of this would be to eliminate the incentive that parents have in preparing, motivating, and teaching their own children, since this would simply result in the child's being penalized for having a good educational background.

Advocates of the second class of definitions argued that the outcomes of education constitute the only true measure of what the schools were doing. They argued that equality according to input resources is compatible with the discredited 'separate but equal' doctrine that governed education in the U.S. South prior to the Brown decision of 1954 in the Supreme Court. They argued that 'provision of equal resources' is too passive a conception of the school's role, placing on the child and the family the full responsibility for taking advantage of those resources. Some children were far better prepared to do so than were others.

We might regard neither of these classes of definitions as fully appropriate, as indicated by the arguments of each against the other, which I have just given. The first class envisions what appears to be a passive a role for schooling, and the second envisions a role for the school that appears unattanable, short of eliminating all family influences whatsoever. An appropriate definition must, it would seem, incorporate some elements of both, with the school taking responsibility not only for a passive provision of resources but also for an intensity of experience that helps to overcome the inequalities of opportunity to which children are subject outside the school.

But even more: These definitional difficulties are symptoms of the fact that schooling can in itself never create full equality of educational opportunity for children, because school is only one portion of the educational influences on children.

No matter how equal these influences, the unequal influences from other institutions, particularly the family, must lead to differing educational opportunities for different children.

It is the failure to recognize this that flaws a work like John Rawl's Theory of Justice, which attempts to lay out a system of institutions that would provide fair equality of opportunity for all. By failing to recognize that inequalities arise not principally from unequal treatment at the hands of central authorities, but from the social structure itself, in the individual households in which different children grow up, Rawl's institutions fail to address the central problems of inequality of opportunity.[1)]

An appropriate conception of formal schooling is one that does recognize the sociale structural sources of unequal opportunities, and sees the school as an institution that in some fashion complements that structure. In this conception, it is not the school itself which provides educational opportunity, but the school in conjunction with existing institutions in the fabric of society, particularly the family.

Once this is recognized, then a further recognition must follow: Since social structures, and in particular families, are very different under different societies, the school must do different things in different societies. A single conception of the school suitable for all social structures is inappropriate. And a single conception of how schools can equalize educational opportunity is equally inappropriate.

Very roughly, three broad phases may be distinguished in the state of a nation's economy and social structure. Parallel to these are three broad phases in the state of a family's economic and social conditions. Thus whatever the phase for a nation as a whole, for example, Phase 2, there will nevertheless be in it some families that are at Phase 1, and some at Phase 3.

I will outline each of these phases, and attempt to give for each a sense of the role of schooling in the equalization of educational opportunity. I distinguish these three phases because in each, the family has a certain set of interests in its children that shape the way it acts toward its children and thus set the environment that the school confronts.

Phase 1: The exploitation of children's labor.

What I will call Phase 1 is an economy in which most households are at or slightly above a subsistence level. An economy based largely on subsistence farming is the most widespread example, though extractive economies in general, in which most occupations are in the primary economic sector, fit this phase, as do village-based societies in which most households are engaged in herding. In such social structures, households directly produce most of what they consume; economic exchange and division of labor are minimal.

In such societies, the labor of children is useful, both because in the diversified activities of the household, there are always tasks that children can carry out, and because the economic level of the household is sufficiently low that the effort of all is needed. Children are not costly to the family because food is ordinarily procuced at home. Families have many children, and exploit their capacity for labor, with little regard for the impact of this upon the children's opportunities. Families have narrow horizons, are inwardly focused, and base little interest in or resources for extending their children's horizons broadly.

In an economic and social structure of this sort, the principal role of the school is in protecting children from exploitation by the family, and in providing a broadening influence beyond the family's horizons. The family constrains and limits the child; the school breaks some of these bonds and reduces the constraints. The school often stands, in such a setting, in an antagonistic position to the family, for the interests of the two often conflict. The school is the liberator of the child from the exploitative grasp of the family.

Yet nations whose economies and social structures are of this sort are the poorest, so that the economic resources necessary to provide educational opportunity are most limited. The nation's capability of providing a strong school system to oppose the constraining force of the family is weakest. Consequently, it is in this phase that educational resources are ordinarily most unequally distributed, between rich and poor villages, or between rich and poor regions. Educational opportunity depends largely on the opportunity provided by the family and the immediately surrounding area.

In this phase, the tangible educational resources, textbooks, teachers, classrooms, libraries, are in short supply. Consequently it is in such societies that these input facilities make most differences in educational outcomes. One can well say that for a nation in Phase 1, equalization of educational opportunity is most dependent on tangible educational resources. In this phase, the first definition of equal educational opportunity, in terms of input resources, is most relevant, since variations in educational opportunity depend most on variations in these resources.

Phase 2: Children as investments for the family.
A post-agricultural, urban, industrial society, engaged largely in manufacturing and some commerce I will call Phase 2. Here the economy is an exchange economy, most labor is performed in full-time jobs, and the family's economic needs are provided mostly through the exchange of wages for goods. Children's labor is no longer needed for the household's economy, and there are fewer possibilities for productive work of children within the household.

In such a society, the family continues to have a strong interest in children, for a more long-range goal. Children are the carriers of the family across generations from the past into future, and investment in children is an investment in human capital for the family's future. A large number of children is no longer valuable for this purpose, but high investments in each one, to increase the status position, economic position, and social respectability of the family in the next generation is.

This change in the family's interest in children has many implications. One is a decline in the birth rate. Another is an increase in the demand for universal education and for equal educational opportunity. The quantity of children is no longer valuable to the family, but the quality of their preparation and training is.

The family is no longer the school's antagonist, but is its most important ally. The family creates a strong motivation for schooling in its children, for the school's goals for the children coincide with the interest of the family.

High academic achievement is to be expected from children whose families are in Phase 2, and high academic achievement in the nation as a whole when the nation is in Phase 2. Family and school are reinforcing each other's actions toward high achievement.

Phase 3: Children as irrelevant.
An advanced industrial society (what Daniel Bell has called a post-industrial society) or a welfare state with a high degree of affluence I will call Phase 3. In this phase, the family's central role in the economy has vanished, and the family itself has become a kind of appendage to the economic structure. It is an institution relevant to consumption, but no longer to production. Its functional role has been reduced to that of childrearing.

The family's central place in the economy and society has been taken over by large corporate bodies - industrial and commercial corporations. As the economic functions of the family are withdrawn to other institutions, the family loses much of its raison d'être, and begins to disintegrate. It is no longer an institution spanning generations, but forms anew with each generation, so the family's interest in children to carry the family into the future declines. The stability of marriages (and thus of households) declines, as the multi-generational family is no longer present to restrain its members from individualistic solutions at the expense of the family.

In such circumstances, we can expect that families would make fewer investments in children, would press strongly toward academic achievement, and would support the goals of the school less completely than in Phase 2. The evidence concerning these actions is not clear. In the United States and some countries in Europe, which are closest to Phase 3, there is an even stronger demand for equal educational opportunity, and more resources invested in education than in the earlier period of Phase 2. But families have shifted much of the responsibility for financing higher education to the government. Parents spend less time with children, and children less time with parents in whole-family settings. Leisure activities instead take place in age-segregated settings: cocktail parties for the adults, rock concerts for the youth. Increasing numbers of children are abandoned, run away from home, or become addicted to drugs, and an increasing number of the children of divorce are unwanted for custody by either mother or father. Yet all these statistics involve a minority of children. At the same time, there is a strong professed interest of parents in their children's educational development.

My own assessment of the trends in the United States is that there is, as one would expect, lesser investment in children than was true thirty or forty years ago, and that the evidence will soon begin to show this more clearly. If I am correct, this means that the school loses much of the active support it had during Phase 2, and that the motivation to achieve which families imparted to their children is less frequent. The school's task, in this condition, comes to be one of supplying not only the resources for learning, but also taking active responsibility for bringing about learning. The school, under these conditions, comes to take over some of the functions which the family once provided, but which it no longer provides.

If this picture is a correct one, it accounts for an otherwise puzzling result: Although in less developed countries there is a strong relationship between the tangible school resources in a region or locality or a school and the level of academic achievement of the students in that region, locality, or school (controlling on family backgrounds of the students), this relationship vanishes or is sharply reduced in highly developed countries. The achievement attributable to the school itself in highly developed countries is almost independent of the level of tangible school resources provided by the community or the nation. The achievement is not independent of the way the school is organized, the disciplinary constraints it imposes on students, and the academic demands it makes on them. But a school with excellent physical resources, laboratories, books, and teacher qualifications, a school with high per pupil expenditures, does not produce high achievement if these less tangible organizational elements are missing.

If the picture I have given is correct, the highly developed counries are moving into Phase 3, in which tangible school resources are in oversupply, not only in the school itself, but in the home, through television, and quite generally throughout the society. The student motivation to learn, which was provided by families in Phase 2, is now problematic. With these tangible resources in oversupply, an increase or decrease of 50% in the school resources does not make much difference in achievement, though it did when these countries were in Phase 1 and Phase 2, and these resources were in short supply.

What is in short supply in the affluent Phase 3 is not these tangible resources, but the motivations that strong families, interested in investing time, effort, and attention in their children, provided in Phase 2. The schools that are most effective in this third phase are those that are able to supply the intangible qualities that impel students to take full advantage of the opportunities provided by the tangible resources. The school, in Phase 3, is one of many elements competing for the attention and interest of children and youth, and what cannot be taken for granted are the motivational forces that direct attention and interest toward school learning, rather than toward the other attractive competitors for this attention and interest.

CONCLUSION

What, then, does provision of educational opportunity consist of? All that I have described above implies that it consists of different things when the social and economic conditions of the nation differ. When a nation is in Phase 1, it consists of the provision of tangible resources for learning, plus legal and other constraints on families' exploitation of their children, so that children are free to take advantage of these resources. Some caution must be introduced here, however, because the mere provision of educational opportunity through formal schooling in an economy with an occupational structure that requires the old skills is harmful to both the child's and nation's future. It has drawn children away from the old skills without being able to make use of the new ones. The activities that were economically helpful to the household were also inculcating certain narrow skills that the child could use as he or she replaced father or mother in the next generation, and the school's influence undercuts the learning of these.

In Phase 2, the nation's task in providing educational opportunity is the simplist: mere provision of the tangible resources of formal schooling. This, combined with the motivation that families - acting in their own interest - provide gives an effective eduational opportunity. And insofar as these resources are provided in different schools with some approximate degree of equality, the nation is providing an effective educational opportunity that is a strong influence in the direction of equal opportunity.

In Phase 3, the school's task in providing educational opportunity becomes more complex, as described earlier, and is no longer satisfied by the provision of tangible school resources. The full scope of the task is unclear, and I suspect that it will be some time before we learn just how it can be best accomplished. The school's role expands, the possibilities for greater equality of opportunity increase as the power of families declines; but the possibilities for educational mediocrity increase as well. Altogether, it is part of a structure of society that is only beginning to unfold, and one about which we have much to learn.

NOTE: 1) For an extended discussion, see Coleman, 'Inequality, Sociology, and Moral Philosophy'. *American Journal of Sociology* 80, No. 3 (November 1974): 739-764.

Phases in social structure and change of educational opportunity, a comment on Coleman's paper.

J. Dronkers
S.I.S.W.O.
Amsterdam
The Netherlands

Comparisons between different societies and their education have always been a promising way in social sciences but at the same moment also a very dangerous one. The danger is the ascription of the emprirically found differences to only one attractive or visible difference in stead of to the whole range of differences between the compared societies and their educational systems. We can see this danger clearly in the three phases prof. Coleman presented to us.
He stresses the different relations between family and their off-springs in the three phases as the main source of the different effects of education and school in different societies and times.
If we believe Coleman's phases for a moment, one can rightly wonder if these family-child relations are indeed so important because there are a lot more differences between farming, industrial and post-industrial societies, which can influence the effects of schooling. The effects of these other differences between the three types of societies can reinforce but also weaken the effect of different family-child relations on schooling and equality.
I will give you an hypothetic example.
Coleman states that in the post-industrial phase the economic role of the family has vanished and been taken over by large corporate bodies. The family begins to disintegrate, makes fewer investments in children, and press strongly toward academic achievement. Let us suppose this picture is basicly correct. An opposite effect of the post-industrial society on schooling can also result from the rise of these large corporate bodies. The bureaycracy and the production processes of these large corporate bodies require more formal schooling than the small factories and shops of the industrial phase. This increased necessity for more education in order to participate in the economic life can fully compensate the less press of the desintegrating family toward academic achievement. Even members of desintegrated families in a post-industrial society will realize the increased importance of educational credentials. This hypothetic example illustrates the danger of comparing societies and times.

I do not say that Coleman's phases are worthless. We need a theoretical underpinning why we want to compare what we compare, and what we expect to find. However, the theoretical description of the compared societies has to be comprehensive in order to be really useful and to avoid political abuse.

This brings me to a second objection against Coleman's phases. His presentation of these phases suggests a kind of evolution theory. This type of theories supposes movements of nations from one phase to another, some early, others late. The movements from one phase to another have to be necessarily followed by changes of educational opportunities. If these phases are usefull, one must find differences in relative educational opportunities between for instance the industrial phase and the post-industrial phase.

The empirical evidence of changing relative educational opportunities of social classes during the transition-period from the industrial to a post-industrial society is not very strong an it supports only partial Coleman's phases. A group Dutch researchers has studied the change of the relative educational opportunities with data of the educational attainment processes of several generations (an english review of this research: Dronkers, 1982). Also other social scientists, especially those who study social mobility, have focussed on the changing mobility-ratios between different societies and times (for a handy review of this research, see Heath, 1981). This research does not show successive phases in relative mobility and educational opportunities, but only found changes in the importance of contrasts between social classes in their relations with education.

Again, I will give an example from our Dutch research on changing educational opportunities. We found the same contrast between the agrarian and the non-agrarian classes in their use of schooling both in the socalled industrial phase and in the post-industrial phase of Dutch society. This cannot be explained by the backwardness of Dutch agrarians or by their exploitation of the labor of their children. The Dutch agrarian sector is one of the most industrialized of the world which does not use children's labor.
A better explanation might be that an agrarian need not to rely on education as one of his important means of this production, in contrast to non-agrarian workers. The great difference between industrial and post-industrial society is the number of people working in this agrarian sector and thus the importance of this contrast to the national educational system.
The same holds for the contrast between stable and non-stable families in their use of schooling both in the industrial phase and in the post-industrial phase. We found that non-stable families had the same low relative educational opportunities in both types of societies (Vrooman and Dronkers, in press). However, the number of non-stable families grew strongly since the end of the last world-war. Therefore the average school is now increasingly confronted with children from non-stable families. Large numbers of non-stable families is however not unique for the post-industrial society. For instance England and France of the eighteenth century had large numbers of desintegrated families and 'irrelevant children' (Jean-Jacques Rousseau).

In other words, there seems not to be successive phases of societies as Coleman supposed but there might be a change of the importance of the contrast between social classes (agrarian versus non-agrarian; manual worker versus brainworker; non-stable families versus stable families; etc.) and therefore a change in their overall impact on the educational system and a change of the relevance of education for those contrasts.

Concluding, I wonder if Coleman's successive phases in the relations between family and schooling is a sound base for the further comparison of societies and generations.

REFERENCES

Dronkers, J.: Do unequal educational changes degrees by legislation, put through by political movements? Paper presented at the 10th world congress of sociology, Mexico-city, 1982.

Heath, A.: *Social Mobility*, Fontana, London, 1981.

Vrooman, C. and J. Dronkers: *A critic on the comparison by Meesters et al.*, in press.

Designing a policy for equality of educational opportunity, a comment on Coleman's paper.

A. Hoogerwerf
Department of Public Administration
Twente University of Technology
The Netherlands

It is a great privilege and pleasure to comment here briefly on a paper of professor James S. Coleman, one of the fathers of research on equality of educational opportunity. I consider his paper on Schooling and Equality [1] as a new and again brilliant contribution of Coleman's thinking on this subject. By placing the problem of inequality of educational opportunity in a comparative and developmental perspective, Coleman has given an important stimulus, not only to research on education, but also to policy research and to the development of public policies. As a political scientist, teaching policy studies in a department of public administration, I will restrict my comments to some remarks from the point of view of policy studies.

The leading thought of Coleman's paper is as far as I can see: " ... the school must do different things in different societies". In other words, the school is looked upon as a policy instrument, that can be used for various purposes, including equality of educational opportunity. The way in which the school must take shape as an instrument for this purpose depends according to Coleman on the conjunction of the school with existing institutions in the fabric of society, particularly the family.

This leading thought is not new, but nonetheless excellent, the more so since it is in agreement with a more general and well founded thought on policies, namely that the effectiveness of a policy depends on the causal relations in the policy field on which the policy is directed. So the choice of a policy instrument should be attuned to those causal relations. This thought was already expressed by Machiavelli (1469 - 1527), who wrote that the attainment of the goal depends on the harmony between the applied means and the disposition of time.[2] Nonetheless this idea is often neglected in public policies.

So far I hope to have made clear that Coleman shows a way towards a scientific contribution to the design of educational policies. My comments will concentrate on this contribution. In doing so, I shall follow very briefly the stages which should characterize, according to my opinion, the process of designing a policy that is as rational and as legitimate as possible.[3] The underlying premise is that a policy should be designed on the basis of reliable, possibly scientific, information and not only on the basis of intuition, experience, political ideology and power.

1. A first phase in the process of designing a policy is the formulation and analysis of the mandate for the designer. The mandate is in this case to design a policy for equality of educational opportunity. An important question is then who are the actors who decide on the acceptance and the implementation of the designed policy, i.e. the policymakers for whom the design has been made. Coleman refers in his paper to various possible actors: the nation, the educational system, and the school. He does not mention the government. His hesitation is understandable. Many governmental policies fail, in the sense that their goals are not attained. A recent survey of Dutch evaluation research on public policies came to the conclusion that the goals of public policies were not attained in 15 of the 17 investigated policy areas[4]. The explanation of this failure lies partly in the fact that the government and its bureaucracy have often too little knowledge of the causal relations in the policy field on which the policy is directed. Governments tend to suffer from autism: they are self-centred and too much taken up with fancies; they live in daydreams in which the connetctions with the outer world are interrupted.
Against this background it is quite understandable to think of the nation, the educational system or the school and not the government as the most appropriate actor for educational policies. Scientifically, however, it is not guaranteed that self-government of the educational system or the school will be more effective with respect to equality of educational opportunity than public administration. This uncertainty is a fascinating challange for education and policy research.

2. A second step in designing a policy is the analysis of the problem on which the policy is directed. Let us assume that the problem is formulated in terms of the American Civil Rights Act of 1964, as a lack of equality of educational opportunity by reason of race, religion, or national origin. Other reasons for inequality, such as cultural minority positions, socio-economic class, place of residence, or sex, may be added. One then still needs a definition of inequality of education. Coleman makes here a useful distinction between two dominant classes of definitions, one having to do with inputs into education, and the other having to do with outputs from the educational system. What I miss here, is a definition in terms of the throughput of the educational system, i.e., a definition of inequality within the educational process proper[5], which has to do with the relations between teachers and pupils and between pupils in the class, and with relations between types of schools in the educational system.

3. As the third phase in designing a policy I consider the formulation of a model of causal relations in the policy field on the basis of theoretical insights and empirical analysis. Coleman chooses a differentiated approach by presenting not one, but three models of the policy field. He defines them as the phases of the exploitation of children's labor, children as investments for the family, and children as irrelevant. This division is very enlightening, but it should be handled carefully. It is more applicable to categories of the population than to societies as a whole. Not only the situation of children as irrelevant (an eye-opening, but horrible term) is characteristic of only part of a population; the same is true for the two other phases. It is also not difficult and not superfluous to think of a fourth phase, that is already reality for part of the population, namely that of children as potentially unemployed persons. Nonetheless, even a fourfold division is incomplete, as the policy field may vary in several respects from time to time, from place to place, and from category to category of the population.

A question that arises here is how the policy field on which the policy is directed should be circumscribed. Coleman speaks of an educational system and educational influences. There are useful theoretical concepts, but I doubt that they are sufficient for the circumscription of the field of a particular educational policy.

What is needed here is a causal model of the policy field. A policy is a system of ends and means. Thinking in terms of ends and means presupposes thinking in terms of causes and effects. So every policy is based implicitely or explicitly on a causal model of the policy field. This causal model is the circumscription of the field on which the policy is directed.

In designing a causal model of a policy field, one should start with the essential dependent variable. This dependent variable can be derived from the mandate for the designer of the policy and from the formulation of the policy problem (stages 1 and 2 of the design process). In this case the essential dependent variable is "lack of equality of educational opportunity". The causal chains which lead to the lack of equality can be presented in a schema of arrows, in which the points indicate the variables, whereas the lines indicate the causal relations between the variables, and the arrows indicate the causal directions. Such a causal model of the policy field should be based as much as possible on scientific theories and empirical research.

Coleman's analysis asks attention for several possible causes of lack of equality of educational opportunity. He mentions in this connextion the stage of economic and social development of society, the class structure, the lack of legal requirements for equalizing educational opportunity, the policies of state and local governments, the place of the family in the social and economic system, the place of the child in the family, or rather the way in which the child is used as an instrument for a selfish policy of the family (a formulation that is cynical and incomplete, but not per se unrealistic), the functions of the school, the library, and the educational system as a whole. Some factors which I miss here, are cultural differences, including language problems, and housing situations.

Also missing is the important difference between variables with a strong explanatory value and manipulable variables. Many of the factors which Coleman mentions, such as the class structure and the place of the child in the family, cannot be manipulated by educational policies. Whereas fundamental research selects independent variables on the basis of the explanatory power of the variables, applied research should also select independent variables[5] on the basis of the possibility to manipulate them by a policy[6]. If we want to diminish in equality of educational opportunity, we should know which of its causes can be influenced. It is my impression that for educational policies these manipulable variables are mainly situated in the school and the educational system as a whole, and not in the family or in the broader structure of society. And even in the educational system the socalled manipulable variables may be tough.

To the manipulable variables belong the educational policies of governments, schools and other institutions in the field of educational politics. So what we need in order to further equality of educational opportunity is not only research on education, but also research on the contents, the processes, and the effects of educational policies of governments and other actors. This means among other things evaluation research and policy experiments in the field of educational policies. It is my conviction that policy evaluation research and policy experiments[7] are still susceptible of extension, but also of improvement, not only in the area of educational policies, but in whichever policy area. I say this with all due respect for Coleman and others who deserve great appreciation for their contributions to evaluation research.

4. A fourth step in the process of designing a policy is the formulation of ultimate goals and evaluation criteria. In doing this the designer should base himself on the mandate, the formulation of the problem and the causal model of the policy field (stages 1, 2 and 3). Besides he should take full account of constraints of a political, juridical, economic, ethical or other nature.

Equality of educational opportunity is, as Coleman[8] has made it clear before, a rather vaguely formulated goal. Many kinds of material and immaterial goods can be distributed unequally according to many kinds of criteria. Inequality of educational opportunity can exist with regard to many goods, such as enrolement as a student, the subject-matter of teaching, the qualifications of the teachers, the support of pupils, the age of textbooks, the school buildings, endownments, grants and so on. Each of these goods can in principle be allocated according to many criteria, such as the capacities of the pupil, his or her interest, social class, race, religion, sex, place of residence, etcetera. Social and political history can be interpreted from this point of vies as a strive for the abolition of irrelevant criteria and the introduction of relevant criteria for the distribution of all kinds of goods.[9]
From this point of view, the formula of "equality of educational opportunity" is not the best possible one. The questions is what are the most relevant criteria. According to my opinion, the most relevant and maybe the only relevant criteria for the allocation of education are the abilities and the interest of the pupil. A good formula for equality in education is then that everybody has a right to receive education according to his or her abilities and interest, unhindered by other factors, such as for instance class, race, religion, sex, place of residence, political conviction and so on. An educational policy attuned to that principle will have the twofold purpose of positively, offering different forms of education according to the abilities and interests of pupils and, negatively, counteracting the possibility that persons from particular categories of the population do not receive education according to their abilities and interest.

Coleman pays also attention to the intriguing question what are the social conditions of the ideal of educational opportunity. He seeks the explanation in the fact that in his phase 2 the quantity of children is no longer valuable to the family, but the quality of their preparation and training is. This explanation may be right, but it is nonetheless somewhat one-sided. Thinking about equality is not only dependent on the social structure, but also on the social and political culture. The principle of equality, which says that equal cases should be dealt with equally, can be specified in many ways. So for instance the Manifest of equals of 1795, stemming from a group around Babeuf, said already in a rather extreme formulation: "As all people have the same needs and the same capacities, let there be for everybody only one and the same schooling"[10]. This specification of the principle of equality should be understood in the context of the social and political culture. The formula of "equality of educational opportunity" is also culture-bound.

5. A fifth step in designing a policy is the formulation of alternative means, which are expected to lead to the attainment of the goal, i.e. to be effective, and which are legitimate as well. Coleman chooses here again a differentiated approach: "A single conception of how schools can equalize educational opportunities is ... inapropriate". In the phase of the exploitation of children's labor (phase 1), the policy instruments consist of the provision of tangible resources for learning, plus legal and other constraints on families' exploitation of their children. In the phase of children as investments for the family (phase 2) the policy instrument is the mere provision of tangible resources for schooling. And in the phase when children are irrelevant (phase 3) the school's task is supplying not only the resources for learning, but also taking active responsibility for bringing about learning.

I think this way of matching of policy instruments and phases in the policy field is incomplete and a bit too tight. In phase 1 the instrument of taking active responsibility by the school may also be effective for instance with regard to children of minority groups. In phases 2 and 3 the legal constraints remain important instruments.

Another objection is that the policy instruments are here formulated very abstractly. A more specific typology of instruments of policies for equality of educational opportunity is necessary. In such a typology more concrete policy instruments should be classified. Such as for instance criteria for enrolment, curricula, teaching materials, the homework system, the length of the school day and the school year, the organization of the school, the qualifications of the teachers, the contacts with the parents, and so on. On the basis of such a typology and research on the effectiveness of the various policy instruments, in due course a theory of educational instruments could be developed. Such a theory would say which instrument or combination of instruments of educational policy would probably be effective for which goal in which situation[11].

6. Designing a policy demands also a comparison of the expected costs and benefits of the application of the alternative policy instruments. The benefits include not only the effectiveness, i.e. the contribution of the means to the attainment of the goal, but also the effects which were not aimed at (side-effects) as far as they are valued positively from the point of view of other goals than those for which they were used. The costs include not only the financial costs, but also all other negatively valued effects of the instruments.

Coleman pays in his paper no attention to the positive side-effects and the costs of instruments for equality of educational opportunity. Both for researchers and for policymakers, however, it is important to know what are the benefits of the diverse forms of equality of educational opportunity for respectively equality in the distribution of income and other goods, social and cultural integration of minority groups, social and political stability, economic development, and employment. It is also important to know what are the costs of the diverse forms of equality of educational opportunity (for instance positive discrimination and the comprehensive school) in terms of the development of more and less gifted pupils, the quality of education (Coleman speaks in passing of educational mediocracy) and public and private financial positions. Another, related problem is the efficiency, i.e. the relation between costs and benefits of variouw instruments of a policy for equal educational opportunity.

7. As next stages in the process of designing a policy I consider the designing of one or more policy models, the designing of the implementation process, and the ultimate formulation of the policy design. Because of limitations of time I will not deal with these stages now. Let it suffice here to say that there is an urgent need for a policy with strives for equality of educational opportunity for children and adults who are potentially or really unemployed persons in a period of an information revolution. Let me add that the succes of policies should not be overestimated.

In summary I hope to have made it clear that Coleman's paper contains various very stimulating contributions to the designing of policies for equality of educational opportunity. These contributions deserve further analysis and elaboration. I have in this connection tried to draw the attention to some possible contributions of education research and policy research to designing a policy which tries to realize the right of everybody to education according to his or her abilities and interest, unhindered by discriminating factors.

NOTES

1. James S. Coleman: *Schooling and Equality*, paper prepared for delivery at IEA General Assembly in Enschede, The Netherlands, August 17, 1983.
2. Niccolo Machiavelli: *Il Principe*, chapter XXV, various editions.
3. A. Hoogerwerf: *Ontwerpen van overheidsbeleid*, syllabus, Enschede, 1983. Cf. Tj. Plomp: *Onderwijskundige technologie: Enige verkenningen*, Enschede, 1982 and W.J. Nijhof: *Over het ontwerpen van curricula*, Enschede, 1983.

4. A. Hoogerwerf (red.): *Succes en falen van overheidsbeleid*, Alphen aan de Rijn, 1983.
5. In the design stage of the research for the Coleman Report, an internal staff memorandum described five types of educational inequality. See James S. Coleman: Problems of Conceptualization and Measurement in Studying Policy Impacts, in: Kenneth M. Dolbeare (ed.): *Public Policy Evaluation*, Bevely Hills/London, 1975, p. 19-40, especially p. 29-30.
6. Robert A. Scott and Arnold R. Shore: *Why Sociology Does Not Apply: A Study for the Use of Sociology in Public Policy*,New York/Oxford, 1979, p. 229-232. Cf. Coleman, 1975, op. cit. p. 25.
7. Some Dutch publications on policy evaluation are: J.Th.A. Bressers: *Beleidseffectiviteit en waterkwaliteitsbeleid: Een bestuurskundig onderzoek*, Enschede, 1983; J.A.M. Maarse: *Uitvoering en effecten van arbeidsmarktbeleid: Een bestuurskundig onderzoek naar het aanvullende werkgelegenheidsbeleid*, Enschede, 1983 and A. Hoogerwerf (ed.), mentioned in note 4. On policy experiments see: F.A. van Vught: *Experimentele beleidsplanning: Bestuurskundige expedities in de jungle van het planningsdenken*, Enschede, 1982.
8. James S. Coleman et al.: *Equality of Educational Opportunity*, Washington D.C. 1966; James S. Coleman: The concept of equality of educational opportunity, in: *Harvard Educational Review, 38 (1)*, Winter, 1968, p. 7-22.
9. A. Hoogerwerf: Gelijkheid en ongelijkheid in het politieke denken, in: J. van den Doel and A. Hoogerwerf (eds.): *Gelijkheid en ongelijkheid in Nederland; Analyse en Beleid*, Alphen aan de Rijn, 1975, p. 15-32.
10. Ph. Buonarrotti: *Conspiration pour l'egalité dite de Babeuf*, Bruxelles, 1982, tomo second, p. 130-136.
11. Cf. L.A. Geelhoed: *De intervenerende staat: Aanzet voor een instrumentenleer*, 's-Gravenhage, 1983 and A. Hoogerwerf: Aanzet voor een instrumentenleer, in: *Tijdschrift voor Openbaar Bestuur*, augustus 1983.

ACKNOWLEDGEMENTS

The author thanks dr. J.Th.A. Bressers, drs. C. Terlouw and prof.dr. E. Warries for comments on a draft for this comment.